Temas Generales para la preparación de la Oposición al Cuerpo Superior de Estadísticos del Estado.

Cuerpo Superior de Estadísticos del Estado
Especialidad de Estadística-Ciencia de Datos.

Almacenamiento y modelos de datos

Tema 3. Álgebra relacional y cálculo relacional.

AUTOR: **Antonio J. Sánchez-Padial**

Asociación Profesional de Cuerpos Superiores de Sistemas y Tecnologías de la Información de las Administraciones Públicas

Creación: Junio 2021

ÍNDICE

1 El Álgebra Relacional

El Álgebra Relacional es un lenguaje teórico construido con una serie de operaciones que trabajan sobre **relaciones**, tal y como se definen en el Modelo Relacional. Todas las operaciones del Álgebra Relacional tienen como resultado una de estas relaciones. Matemáticamente, esto significa que el conjunto de las relaciones está **cerrado** sobre estas operaciones. La consecuencia más notable de esta propiedad es que los resultados de las operaciones de álgebra relacional pueden ser los operandos de otras, permitiendo el anidamiento de operaciones y con esto una gran expresividad.

Cada operación del álgebra relacional, como lenguaje teórico, opera sobre todas las tuplas de las relaciones involucradas a un tiempo, sin necesidad de bucles.

En este capítulo se recogen las ocho operaciones fundamentales propuestas por Codd (1972a), aunque se han desarrollado algunas otras:

- Selección
- Proyección
- Renombre
- Unión
- Intersección
- Sustracción
- Producto cartesiano
- *Join* (o reunión)
- División

Frecuentemente se clasifican estas operaciones en base a distintos criterios:

Según el número de relaciones sobre las que opera, distinguimos entre operaciones **unarias** (Selección, Proyección, Renombre), que operan sobre una única relación, y **binarias** (Unión, Intersección, Sustracción, Producto cartesiano, *Join* y División) que operan sobre dos relaciones.

También es posible distinguir entre operaciones **básicas** (Selección, Proyección, Renombre, Unión y Sustracción) y operaciones **derivadas** (Intersección, *Join* y División) según si es posible o no definirlas en base a otras operaciones del álgebra relacional.

También se incluyen las operaciones de Asignación y Renombrado que facilitan la simplificación de las expresiones del Álgebra Relacional.

En las siguientes definiciones usaremos las relaciones R y S, que se definen sobre los atributos A = $\{ a_1, a_2, \ldots, a_N \}$ y B = $\{ b_1, b_2, \ldots, b_M \}$, respectivamente.

○ Operaciones unarias

Como hemos adelantado existen tres operaciones unarias: la Selección, la Proyección y el Renombre.

▪ Selección

La Selección (σ) opera sobre una única relación R, su resultado es una relación que contiene solo aquellas tuplas de R que satisfacen la condición especificada o predicado. Se utiliza con la siguiente expresión:

$$\sigma_{predicado}(R)$$

- Ejemplo. Operación Selección

Muestras las provincias con más de 1 000 000 de habitantes.

$\sigma_{habitantes > 1000000}(Provincias)$

En este ejemplo, la operación de selección se aplica a la relación Provincias. El resultado serán todas las tuplas de la relación en los que se cumpla el predicado habitantes > 1 000 000.

Provincia	Comunidad Autónoma	Capital	Habitantes	Costa
Asturias	Principado de Asturias	Oviedo / Uviéu	1 018 784	Atlántica
Balears, Illes	Illes Balears	Palma	1 171 543	Mediterránea
Barcelona	Cataluña	Barcelona	5 743 402	Mediterránea
Bizkaia	Euskadi	Bilbao	1 159 443	Atlántica
Cádiz	Andalucía	Cádiz	1 244 049	Atlántica[1]
Coruña, A	Galicia	A Coruña	1 121 815	Atlántica
Madrid	Comunidad de Madrid	Madrid	6 779 888	*Null*
Málaga	Andalucía	Málaga	1 685 920	Mediterránea
Murcia	Región de Murcia	Murcia	1 511 251	Mediterránea
Palmas, Las	Canarias	Las Palmas de Gran Canarias	1 131 065	Atlántica
Santa Cruz de Tenerife	Canarias	Santa Cruz de Tenerife	1 044 887	Atlántica
Sevilla	Andalucía	Sevilla	1 950 219	*null*
Valencia / València	Comunidad Valenciana	València	2 591 875	Mediterránea

Nótese que es posible relacionar selecciones más complejas utilizando operadores lógicos ∧ (Y), ∨ (O), ~ (NO) y sus combinaciones en el predicado.

- ## Proyección

Como la Selección, la Proyección es una operación unaria, se aplica a una única relación R. La Proyección escoge un subconjunto de los atributos de R (a_1, a_2, …, a_k), y devuelve una relación cuyas tuplas contienen únicamente esos valores. Puesto que la relación resultado es un conjunto, se eliminan los duplicados. Se expresa con la siguiente operación:

$$\Pi_{a_1,a_2,\ldots,a_k}(R)$$

- Ejemplo. Operación Proyección

Lista de las comunidades autónomas con las costas que bañan sus provincias

$$\Pi_{Comunidad\ Autónoma,Costa}(Provincias)$$

En este ejemplo la proyección genera una relación cuyos atributos son los atributos indicados de la relación Provincias: Comunidad Autónoma y Costa; incluyendo todos los valores de esos atributos en las tuplas de la relación Provincias, eliminando cualquier repetición. Por ejemplo, las provincias de la Comunidad Valenciana producirían tres tuplas idénticas *(Comunidad Valenciana, Mediterránea)* de las que se suprimen las duplicadas.

Comunidad Autónoma	Costa
Castilla La Mancha	*null*
Comunidad Valenciana	Mediterránea
Andalucía	Mediterránea
Euskadi	*null*
Asturias	Atlántica
Extremadura	*null*
Illes Balears	Mediterránea
Cataluña	Mediterránea
Euskadi	Atlántica
Castilla y León	*null*
Andalucía	Atlántica

[1] Es bien sabido que Cádiz tiene costa atlántica y mediterránea. Se incluye únicamente el valor *Atlántica* para simplificar el modelo.

Cantabria	Atlántica
Andalucía	*null*
Galicia	Atlántica
Cataluña	*null*
Comunidad de Madrid	*null*
Región de Murcia	Mediterránea

Comunidad Foral de Navarra	*null*
Galicia	*null*
Canarias	Atlántica
La Rioja	*null*
Aragón	*null*

Operaciones de conjunto

Puesto que las relaciones del Modelo Relacional se definen como conjuntos, es posible realizar sobre ellas las operaciones clásicas de conjuntos: Unión, Sustracción, Intersección y Producto cartesiano.

Como veremos la Unión, la Sustracción y el Producto cartesiano se consideran operaciones básicas, mientras que la Intersección es una operación derivada.

Unión

En teoría de conjuntos sería posible considerar la unión de cualquier par de relaciones R y S. Sin embargo, para que el resultado de esta unión sea una relación del modelo relacional, es necesario que R y S sean **compatibles para la unión**.

Dos relaciones R y S son **compatibles para la unión** si tienen el mismo número de atributos, y para cada pareja ordenada de atributos de R y S tienen el mismo dominio. El nombre de los atributos puede ser distinto en las dos tablas, no afecta su compatibilidad.

Para dos relaciones R y S que sean compatibles para la unión, el resultado de la unión consiste en todas las tuplas de R seguidas de todas las tuplas de S, eliminando los duplicados. Los nombres de los atributos de la relación resultado son los de los atributos del primer operando (R en nuestro ejemplo). Se expresa con la simbología bien conocida de teoría de conjuntos:

$$R \cup S$$

Es frecuente utilizar la Proyección como paso previo a la Unión, para forzar la compatibilidad para la unión de los operandos.

- Ejemplo. Operación de Unión

Listado de todas las capitales de provincia y de comunidad autónoma.

$$\Pi_{Capital}(Provincias) \cup \Pi_{Capital}(Comunidades\ Autónomas)$$

Antes de proceder a la Unión es necesario realizar la Proyección de ambas relaciones a los atributos de nuestro interés. De otro modo la unión no tendría sentido en el álgebra relacional.

En el resultado se eliminarían las repeticiones, aquellas ciudades que son a la vez capitales de una provincia y de una comunidad autónoma, como Valladolid o Logroño.

Sustracción

La operación de Sustracción define una relación con todas las tuplas de R, que no estén en S. Los esquemas de R y S tienen que ser compatibles para la unión, para que el resultado tenga sentido en el Álgebra Relacional. Se expresa con el signo de la resta.

$$R - S$$

Ejemplo. Operación de Sustracción

Listado de las capitales de comunidad autónoma que no son capitales de provincia.

$$\Pi_{Capital}(Comunidades\ Autónomas) - \Pi_{Capital}(Provincias)$$

Como en el caso anterior es necesario realizar sendas proyecciones para provocar la compatibilidad para la unión de las relaciones que participan en la Sustracción.

Capital
Santiago de Compostela
Mérida

Intersección

La operación de Intersección define una relación consistente en el conjunto de todas las tuplas que se encuentran tanto en R como en S. Como en la Unión y la Sustracción, los operandos deben ser compatibles para la unión, para que el resultado de la operación tenga sentido en el Álgebra Relacional. Se representa con la notación bien conocida de teoría de conjuntos:

$$R \cap S$$

La Intersección es una operación derivada ya que se puede definir como una combinación de otras operaciones básicas del álgebra relacional:

$$R \cap S \equiv R - (R - S)$$

Ejemplo. Operación de Intersección

Listado de las capitales de provincia que son capitales de comunidad autónoma

$$\Pi_{Capital}(Comunidades\ Autónomas) \cap \Pi_{Capital}(Provincias)$$

Capital
Sevilla
Zaragoza
Oviedo / Uviéu
Palma
Las Palmas de Gran Canaria
Santa Cruz de Tenerife
Santander
Valladolid
Toledo
Barcelona
València

Madrid	Vitoria-Gasteiz
Murcia	Logroño
Pamplona	

Producto cartesiano

El Producto cartesiano es un viejo conocido de los estudiantes de secundaria. La operación de Producto cartesiano define una relación que incluye la concatenación de todas las tuplas de R con todas las tuplas de S. Se representa de la siguiente manera:

$$R \times S$$

La operación de Producto cartesiano multiplica dos relaciones para obtener una nueva relación que consiste en todas las posibles parejas de tuplas de las dos relaciones. Así, si una relación tiene I tuplas y N atributos, y la otra tiene J tuplas y M atributos, la relación resultado de su Producto cartesiano tendrá (I * J) tuplas con (N + M) atributos.

Cuando ambas relaciones operando tienen atributos con el mismo nombre, estos atributos pasan a la relación resultado añadiendo como prefijo a su nombre el de la relación de la que provienen con un punto de separación. De este modo se mantiene la norma del Modelo relacional que impide que una relación tenga múltiples atributos con el mismo nombre.

- Ejemplo. Operación de Producto cartesiano

Mostrar la lista de presidentes del gobierno de España desde 1977, con el nombre y fecha de creación de su partido

Los datos relativos a los presidentes del gobierno se encuentran en la tabla Presidentes, y los relativos a los partidos políticos, en la tabla Partidos. Para mostrar cada presidente del gobierno con la información de su partido será necesario combinar las dos tablas:

$$\Pi_{Nombre,Siglas}(Presidentes) \times \Pi_{Siglas,Nombre,FechaCreación}(Partidos)$$

Nombre	Siglas
Adolfo Suárez	UCD
Leopoldo Calvo Sotelo	UCD
Felipe González	PSOE
José María Aznar	PP
José Luis Rodríguez Zapatero	PSOE
Mariano Rajoy	PP
Pedro Sánchez	PSOE

Proyección de la relación Presidentes: $\Pi_{Nombre,Siglas}(Presidentes)$

Siglas	Nombre	Fecha Creación
UCD	Unión de Centro Democrático	1977
PSOE	Partido Socialista Obrero Español	1879
PP	Partido Popular	1989[2]

Proyección de la relación Partidos: $\Pi_{Siglas,Nombre,FechaCreación}(Partidos)$

[2] En 1977 si se considera su fundación como Alianza Popular

Presidente.Nombre	Presidentes .Siglas	Partidos. Siglas	Partido.Nombre	Fecha Creación
Adolfo Suárez	UCD	UCD	Unión de Centro Democrático	1977
Adolfo Suárez	UCD	PSOE	Partido Socialista Obrero Español	1879
Adolfo Suárez	UCD	PP	Partido Popular	1989
Leopoldo Calvo Sotelo	UCD	UCD	Unión de Centro Democrático	1977
Leopoldo Calvo Sotelo	UCD	PSOE	Partido Socialista Obrero Español	1879
Leopoldo Calvo Sotelo	UCD	PP	Partido Popular	1989
Felipe González	PSOE	UCD	Unión de Centro Democrático	1977
Felipe González	PSOE	PSOE	Partido Socialista Obrero Español	1879
Felipe González	PSOE	PP	Partido Popular	1989
José María Aznar	PP	UCD	Unión de Centro Democrático	1977
José María Aznar	PP	PSOE	Partido Socialista Obrero Español	1879
José María Aznar	PP	PP	Partido Popular	1989
José Luis Rodríguez Zapatero	PSOE	UCD	Unión de Centro Democrático	1977
José Luis Rodríguez Zapatero	PSOE	PSOE	Partido Socialista Obrero Español	1879
José Luis Rodríguez Zapatero	PSOE	PP	Partido Popular	1989
Mariano Rajoy	PP	UCD	Unión de Centro Democrático	1977
Mariano Rajoy	PP	PSOE	Partido Socialista Obrero Español	1879
Mariano Rajoy	PP	PP	Partido Popular	1989
Pedro Sánchez	PSOE	UCD	Unión de Centro Democrático	1977
Pedro Sánchez	PSOE	PSOE	Partido Socialista Obrero Español	1879
Pedro Sánchez	PSOE	PP	Partido Popular	1989

Producto cartesiano de las proyecciones de las relaciones Presidentes y Partidos

La tabla sobre este párrafo muestra el resultado de este Producto cartesiano. Se puede ver que muestra más información de la necesaria, con tuplas que incluyen información que no corresponde con la realidad. Verdaderamente, el Producto cartesiano devuelve todas las combinaciones posibles entre las tuplas de dos relaciones. Cuando, como en este caso, se espera que solo aparezcan las tuplas que corresponden a un enlace real entre los elementos de las relaciones es necesario aplicar una operación de Selección al resultado:

$$\sigma_{Presidentes.Siglas=Partidos.Siglas}(\Pi_{Nombre,Siglas}(Presidentes) \times \Pi_{Siglas,Nombre,FechaCreación}(Partidos))$$

Presidente.Nombre	Presidentes .Siglas	Partidos. Siglas	Partido.Nombre	Fecha Creación
Adolfo Suárez	UCD	UCD	Unión de Centro Democrático	1977
Leopoldo Calvo Sotelo	UCD	UCD	Unión de Centro Democrático	1977
Felipe González	PSOE	PSOE	Partido Socialista Obrero Español	1879
José María Aznar	PP	PP	Partido Popular	1989
José Luis Rodríguez Zapatero	PSOE	PSOE	Partido Socialista Obrero Español	1879
Mariano Rajoy	PP	PP	Partido Popular	1989
Pedro Sánchez	PSOE	PSOE	Partido Socialista Obrero Español	1879

Producto cartesiano restringido de las proyecciones de las relaciones Presidentes y Partidos

▪ Operador de Asignación

Las operaciones de álgebra relacional pueden ser de enorme complejidad. El anidamiento en múltiples niveles y la yuxtaposición de operaciones puede hacer difícil su lectura. Para su simplificación es posible usar el operador asignación ←. El operador asignación ← funciona de modo semejante a la asignación en un lenguaje de programación (`var a = 2`)

Se representa como:

$$R \leftarrow E$$

E es una expresión de Álgebra Relacional; normalmente, una combinaciones de operaciones de sobre una o varias relaciones. Recordemos que el resultado de las operaciones de Álgebra Relacional es siempre otra relación. A partir de esta operación de Asignación, la relación R tendrá la misma estructura que E: mismo número de atributos, con los mismos nombres y tipos, y también contendrá las mismas tuplas. Será una copia exacta.

Podríamos simplificar el ejemplo anterior del producto cartesiano de la siguiente manera:

$R1 \leftarrow \Pi_{Nombre,Siglas}(Presidentes)$

$R2 \leftarrow \Pi_{Siglas,Nombre,FechaCreación}(Partidos)$

$R3 \leftarrow R1 \times R2$

$Resultado \leftarrow \sigma_{Presidentes.Siglas=Partidos.Siglas}(R3)$

▪ Operador de Renombrado

El operador Renombrado nos permite simplificar expresiones de un modo similar a como lo hace el operador de Asignación. El operador de renombrado tiene la siguiente forma:

$$\rho_{R(a_1,a_2,\dots\, a_n)}(E)$$

Como resultado de la operación la relación S será renombrada como R. a_1, a_2, ... a_n es una lista de nombres de atributos, que se asignarán como nombres de los atributos de R en lugar de los que tuviera E. Naturalmente, n debe coincidir con el número de atributos de E. Si se omite la lista de atributos a_1, a_2, ... a_n la siguiente equivalencia es cierta:

$$\rho_R(E) \equiv R \leftarrow E$$

○ Operaciones Join (reunión)

Como hemos visto en el ejemplo, el producto cartesiano devuelve todas las combinaciones posibles entre las tuplas de dos relaciones, sin tener en cuenta que pueda o no existir un vínculo real entre ambas.Por tanto, como en el ejemplo, es muy frecuente combinar las operaciones de Producto cartesiano y Selección para obtener un resultado acorde a la realidad. La operación *Join* (reunión, en algunas traducciones al castellano con poco arraigo) es la combinación de estas operaciones.

Existe toda una familia de operaciones *Join*. La forma más básica es la siguiente (obsérvese la equivalencia con la combinación de Selección y Producto cartesiano):

$$R \bowtie_\theta S \equiv \sigma_\theta(R \times S)$$

Θ es un predicado lógico, análogo al que utiliza la operación de Selección. Debido al uso de la letra griega Θ (theta) para identificar este predicado esta forma básica del *Join* también es denominada Theta *Join*, *Join* Theta o Θ-Join.

El *Join* es una de las operaciones más difíciles de implementar de modo eficiente en un Sistema de Gestión de Bases de Datos (SGBD) y es una de las razones por las que los sistemas relacionales tienen problemas intrínsecos de rendimiento.

Además de este *Join* (básico o theta), existen algunas variaciones que resultan útiles en determinadas ocasiones:

- El Equijoin (un tipo particular de Join Theta)
- Natural Join
- Outer Join (o Join externo)
- Semijoin

▪ Equijoin

En el caso que el predicado Θ contenga únicamente igualdades (=), se le denomina **Equijoin.** Es el caso más habitual de *Join*. Las igualdades suelen ser emparejamientos de atributos, frecuentemente claves y claves externas, de las relaciones operando.

- Ejemplo. Operación Equijoin

Mostrar la lista de presidentes del gobierno de España desde 1977, con el nombre y fecha de creación de su partido.

Anteriormente, usamos el Producto cartesiano para resolver este ejemplo, aunque fue necesario combinarlo con una operación de Selección. Podemos resolverlo de un modo más compacto con un Equijoin:

$$(\Pi_{Nombre,Siglas}(Presidentes)) \bowtie_{Presidentes.Siglas=Partidos.Siglas} (\Pi_{Siglas,Nombre,FechaCreación}(Partidos))$$

O

$$R \leftarrow (\Pi_{Nombre,Siglas}(Presidentes)) \bowtie_{Presidentes.Siglas=Partidos.Siglas} (\Pi_{Siglas,Nombre,FechaCreación}(Partidos))$$

El resultado es idéntico al del ejemplo visto para el Producto cartesiano.

▪ *Join* natural (*Natural join*)

Hemos visto como el *Join* simplifica la expresión de la combinación de un Producto cartesiano y una Selección. De un modo similar, el *Join* natural (o *natural join* en inglés) simplifica uno de los casos de *Join* más frecuentes. Sean R y S son dos relaciones que tienen n atributos (a1, a2, … an) en común - es decir, con el mismo nombre y tipo. El Join natural es un Equijoin entre estas relaciones R y S sobre todos estos atributos comunes x. Se expresa de la siguiente manera:

$$R \bowtie S$$

Además, el resultado del *Join* natural será una relación que contendrá solo una copia de los atributos comunes. Por este motivo las siguientes expresiones no son equivalentes:

$$R \bowtie S \not\equiv R \bowtie_{R.a_1=S.a_1 \wedge \ldots \wedge R.a_n=S.a_n} S$$

La expresión de la derecha conserva dos copias de los atributos comunes, cada una con el prefijo de su relación de origen (R.a_1, S.a_1, …). La expresión de la izquierda contiene una única copia, que, por tanto, no necesita ningún prefijo.

- Ejemplo. Operación Join natural

Una vez más vamos a mostrar el listado de presidentes del gobierno junto a la información sobre los partidos en los que militaban.

$$(\Pi_{Nombre,Siglas}(Presidentes)) \bowtie \rho_{P(Siglas,Partido,FechaCreación)}(\Pi_{Siglas,Nombre,FechaCreación}(Partidos))$$

Nombre	Siglas	Partido	FechaCreación
Adolfo Suárez	UCD	Unión de Centro Democrático	1977
Leopoldo Calvo Sotelo	UCD	Unión de Centro Democrático	1977
Felipe González	PSOE	Partido Socialista Obrero Español	1879
José María Aznar	PP	Partido Popular	1989
José Luis Rodríguez Zapatero	PSOE	Partido Socialista Obrero Español	1879
Mariano Rajoy	PP	Partido Popular	1989
Pedro Sánchez	PSOE	Partido Socialista Obrero Español	1879

Producto cartesiano restringido de las proyecciones de las relaciones Presidentes y Partidos

Es conveniente analizar algunos detalles de este ejemplo. En primer lugar, y al contrario que cuando resolvimos este ejemplo con la operación Equijoin, solo aparece una vez el atributo Siglas. El *Join* natural elimina el atributo duplicado.

También es conveniente fijarse en como se usa el operador Renombrado en el operando derecho del *Join* natural. Su propósito es renombrar el atributo Nombre que pasa a llamarse Partido. Si no lo renombráramos el atributo Nombre de Partidos sería usado para enlazar con el atributo Nombre de Presidentes, obteniéndose una relación vacía en el resultado al no haber ninguna coincidencia entre esos campos.

▪ *Join* externo (*Outer join*)

En el resultado de un *join* solo aparecen aquellas tuplas que tienen un vínculo, a través de sus atributos, con las tuplas de la otra relación. Los *Join* externos (*outer join*, en inglés) devuelven tanto las tuplas resultado de un *Join* básico o natural junto al resto de tuplas que no comparten dicho vínculo. Por contraste con estos *Join* externos, podemos usar el denominador internos para referirnos a los vistos anteriormente.

El *Join* externo puede ser izquierdo, derecho o completo (*left outer join, right outer join* y *full outer join*). El *Join* externo izquierdo devuelve tanto las tuplas que devolvería un *Join* interno, junto a todas las tuplas de la relación en el lado izquierdo del *Join* que no cumplen la condición para emparejarse con alguna del lado derecho. En dichas tuplas, los atributos que vienen de la relación en la parte derecha del *Join* tendrán valor nulo (*null*). Se simboliza con ⟕ y normalmente se abrevia como Join izquierdo.

El *Join* externo derecho es simétrico al anterior, desde el punto de vista de la relación en el lado derecho del *Join.* Su símbolo es ⟖

El *Join* externo completo, devuelve el resultado del Join interno, junto a las tuplas no emparejadas que devolverían tanto el *Join* externo izquierdo como el derecho. Su símbolo es ⟗, y se abrevia como Join completo.

Los Join externos pueden ser naturales o theta según si incluyen o no el predicado a cumplir como subíndice del operador.

El *Join* externo está cada vez más disponible en los sistemas relacionales y forma parte del estándar de SQL. La ventaja de un *Join* externo es que se conserva la información, es decir, se conservan tuplas en el resultado que se perderían con otro tipo de *Join*.

- Ejemplo. Join externo izquierdo

Listado de las Comunidades Autónomas con sus kilómetros de costa

Supongamos que tenemos una relación ComunidadesAutonomas y otra CostaComunidadesAutonomas. En esta segunda, solo aparecen las comunidades autónomas con costa, junto a la información sobre esta.

ComunidadAutonoma	Costa	Kilómetros
Galicia	Atlantica	1629
Canarias	Atlántica	1501
Illes Balears	Mediterránea	1283
Andalucía	Mediterránea	573
Andalucía	Atlántica	337
Cataluña	Mediterránea	699
Comunidad Valenciana	Mediterránea	518
Principado de Asturias	Atlántica	401
Cantabria	Atlántica	284
Región de Murcia	Mediterránea	274
Euskadi	Atlántica	246

Relación de datos costeros de las comunidades autónomas (CostaComunidadesAutónomas)

Si queremos mostrar un listado de todas las Comunidades Autónomas con esta información, tanto si tienen costa como si no, tendremos que hacer uso de un *Join* izquierdo, de la siguiente manera:

$$\Pi_{ComunidadAutonoma}(ComunidadesAutonomas) ⟕ CostaComunidadesAutonomas$$

ComunidadAutonoma	Costa	Kilómetros
Galicia	Atlantica	1629
Canarias	Atlántica	1501
Illes Balears	Mediterránea	1283
Andalucía	Mediterránea	573
Andalucía	Atlántica	337
Cataluña	Mediterránea	699
Comunidad Valenciana	Mediterránea	518
Principado de Asturias	Atlántica	401
Cantabria	Atlántica	284
Región de Murcia	Mediterránea	274
Euskadi	Atlántica	246
Castilla y León	*null*	*null*
La Rioja	*null*	*null*
Comunidad Foral de Navarra	*null*	*null*
Aragón	*null*	*null*
Extremadura	*null*	*null*
Comunidad de Madrid	*null*	*null*
Castilla – La Mancha	*null*	*null*

Resultado del *Join* izquierdo entre ComunidadesAutonomas y CostaComunidadesAutonomas

- **Semijoin**

La operación Semijoin define una relación que contiene las tuplas de R que participan en el Join entre R y S y cumplen el predicado F.

$$R \rhd_F S$$

La operación Semijoin realiza un Join theta de las dos relaciones y luego proyecta sobre los atributos del primer operando. Una ventaja del Semijoin es que maneja menos tuplas para formar el join. Es particularmente útil para calcular joins en sistemas distribuidos. El Semijoin puede ser escrito en términos de las operaciones Join y Proyección:

$$R \rhd_F S \equiv \Pi_A(R \bowtie_F S)$$

Donde A es el conjunto de atributos de R. Si se omite el predicado F, el semijoin se comporta como la combinación de una proyección sobre los atributos de R y un *Join* natural.

- **Ejemplo. Operación Semijoin**

Listado de las capitales de Comunidad Autónoma con mar.

La información solicitada (las capitales) está en la relación de ComunidadesAutónomas, pero la selección se hace por la participación en la relación de CostaComunidadesAutónomas. Podemos resolverlo con un Semijoin de la siguiente manera:

$$\Pi_{Capital}(ComunidadesAutonomas \ \rhd_{Nombre=ComunidadAutonoma} \ CostaComunidadesAutonomas)$$

Capital
Santiago de Compostela
Las Palmas de Gran Canarias
Palma
Sevilla
Barcelona
València
Oviedo
Santander
Murcia
Vitoria - Gasteiz

Operación División

Si S es una relación cuyos atributos son un subconjunto de los de R, la operación de División devuelve una relación sobre los atributos no comunes (C) con las tuplas de R que se pueden emparejar con todas las tuplas de S.

$$R \div S$$

Es posible expresar la División en términos de las operaciones básicas:

$T_1 \leftarrow \Pi_C(R)$

$$T_2 \leftarrow \Pi_C\big((T_1 \times S) - R\big)$$
$$T \leftarrow T_1 - T_2$$

- **Ejemplo. La operación División**

Obtener la lista de las comunidades autónomas que con salida a los dos mares.

$$\Pi_{ComunidadAutonoma,Costa}(Provincias) \div \Pi_{Costa}(Provincias)$$

Comunidad Autónoma
Andalucía

Resultado de la División entre las Comunidades Autónomas con su costa y las costas de la Península

Operaciones agregadas y agrupamiento

Hasta ahora hemos visto como recuperar y componer tuplas a partir de las tuplas, atributos y valores de una o más relaciones. Con las operaciones **agregadas**, es posible obtener resultados agregados como sumatorios, similares a los totales que aparecerían en la parte inferior de un informe. Las operaciones de **agrupamiento** nos permiten seleccionar algunas de las tuplas antes de realizar una operación agregada, de un modo similar a los subtotales de un informe.

Estas operaciones no pueden ser realizadas combinando las operaciones básicas del álgebra relacional que hemos visto hasta ahora.

Operaciones de agregación

Aplica la lista de funciones agregadas, AL, a la relación R, para obtener una relación sobre la lista agregada. AL contendrá una o más parejas (<función_agregada>, <atributo>).

$$\Im_{AL}(R)$$

Las principales funciones agregadas son:

- COUNT – devuelve el número de valores en el atributo asociado.
- SUM – devuelve la suma de los valores en el atributo asociado.
- AVG – devuelve la media de los valores en el atributo asociado.
- MIN – devuelve el valor más pequeño en el atributo asociado.
- MAX – devuelve el valor más alto en el atributo asociado.

- **Ejemplo. Operaciones agregadas**

a) ¿Cuántas provincias tienen más de 1 000 000 habitantes?

Podemos encontrar el resultado con la función agregada COUNT:

$$\rho_{R(NoProvincias)}\big(\Im_{COUNT\ provincia}(\sigma_{habitantes>1\,000\,000}(Provincias))\big)$$

NoProvincias
13

b) Encuentra el mínimo, el máximo y la media de habitantes de las provincias de España

$$\rho_{R(min,max,media)}\big(\Im_{MIN\ habitantes,MAX\ habitantes,AVG\ habitantes}(Provincias)\big)$$

Min	Max	Media
88 884	6 779 888	945 590

▪ Operación de agrupamiento

Agrupa las tuplas de la relación R sobre los atributos de agrupamiento, GA, y a continuación aplica la lista de funciones agregadas AL para definir la relación resultante. AL contiene una o más parejas (<función agregada>, <atributo>). La relación resultado contiene los atributos de agrupamiento, GA; junto con los resultados de cada una de las funciones agregadas.

$$_{GA}\Im_{AL}(R)$$

La forma general de la operación de agrupamiento es la siguiente:

$$_{a_1,a_2,\ldots\, a_n}\Im_{\langle A_p a_p\rangle,\langle A_q a_q\rangle,\ldots\,\langle A_z a_z\rangle}\;(R)$$

donde R es una relación, a_1, a_2, …, a_n son atributos de R sobre los que haremos grupos, a_p, a_q, …, a_z son otros atributos de R, y A_p, A_q, …. A_z son funciones agregadas. Las tuplas de R se dividen en grupos tales que:

- Todas las tuplas en el mismo grupo tienen los mismos valores para a_1, a_2, …, a_n;
- Las tuplas de distintos grupos tienen distintos valores para a_1, a_2, …, a_n.

- Ejemplo. Operación de agrupamiento

Encuentra el número de comunidades autónomas y el número de kilómetros que tiene cada costa.

$$\rho_{R(Costa,numCCAA,kmCosta)}\left({}_{costa}\Im_{COUNT\ ComunidadAutonoma,SUM\ kil\acute{o}metros}(CostaComunidadesAutonomas)\right)$$

Costa	numCCAA	kmCosta
Atlántica	6	4398
Mediterránea	5	3347

2 El Cálculo Relacional

Al contrario de lo que su nombra pueda insinuar, el Cálculo Relacional no está relacionado con el Cálculo infinitesimal de las matemáticas, si no con el **cálculo de predicados**, una rama de la lógica simbólica. Existen dos formas de Cálculo relacional: el Cálculo relacional de **tuplas**, según la propuesta original de Codd (1970), y el cálculo relacional de **dominios**, propuesto por Lacroix y Pirotte (1977).

En cálculo de predicados, un **predicado** es una función con argumentos a la que se puede asignar un valor de verdad. Cuando asignamos valores a estos argumentos, la función genera una expresión que denominamos **proposición**, de la que podemos decir que es verdadera o falsa. Por ejemplo: "Sevilla es una provincia" o "Teruel tiene más habitantes que Madrid" son proposiciones, ya que podemos determinar si son verdaderas o falsas. En el primer caso, tenemos una función "ser una provincia ", con un argumento (Sevilla), en la segunda tenemos la función "tiene más habitantes que", con dos argumentos (Teruel y Madrid).

Además de valores concretos, o en su lugar, los predicados pueden llevar variables: como "x es una provincia". Cada variable debe tener un **dominio** asociado. El valor de verdad de la proposición puede cambiar, cuando sus variables toman unos u otros valores dentro sus dominios. Por ejemplo, si el dominio de *x* es el conjunto de todas las ciudades de España y hacemos que *x* tome el valor Zaragoza, la proposición "Zaragoza es una capital de provincia" será cierta. Si *x* tomará el nombre de otra ciudad que no fuera capital de provincia (como Gijón), la proposición sería falsa.

Si P es un predicado, el conjunto de todas los valores *x* para los que P es cierto se escribe:

$$\{x \mid P(x)\}$$

Es posible crear predicados compuestos mediante la composición de predicados simples con los conectores lógicos ∧ (Y), ∨ (O), ~ (NO)

o Cálculo Relacional de Tuplas

El objetivo del Cálculo relacional de tuplas es encontrar las tuplas para las que determinado predicado es verdadero. Para ello hace uso de **variables de tupla.** Una variable de tupla es una variable cuyo ámbito es una relación, es decir, una variable que solo puede tomar valores en las tuplas de esa relación. Cuando la variable P pertenece al ámbito de la relación Provincias, la siguiente expresión es cierta, y falsa, en caso contrario:

$$Provincias(P)$$

Para indicar la búsqueda "Encuentre todas las tuplas S, para las que F(S) es cierto)

$$\{S \mid F(S)\}$$

F es una **fórmula** (**fórmula bien formada** en lógica matemática). Por ejemplo, para expresar la búsqueda " Encuentra las provincias con más de 1 000 000 habitantes", escribiremos:

$$\{P \mid Provincias(P) \land P.habitantes > 1000000\}$$

Para encontrar una tupla formada por algunos atributos en concreto, escribiremos:

$$\{P.nombre, P.habitantes \mid Provincias(P) \land P.habitantes > 1000000\}$$

▪ Cuantificadores

Los cuantificadores son un elemento del cálculo proposicional que permite aumentar la expresividad de las proposiciones. Por un lado, tenemos el **cuantificador existencial** ∃ (leído "existe"). Se usa en fórmulas que deben ser ciertas al menos para un ejemplo, como en:

$$\{Provincias(P) \wedge (\exists Q)(\ Presidentes(Q) \wedge Q.provinciaNacimiento = P.nombre)\}$$

Que significa, "existe un tupla de Presidentes cuya provinciaNacimiento coincide con el nombre de la tuplas P de la relación Provincias". Por su parte, el **cuantificador universal** ∀ (leído "para todo") se usa en fórmulas que deben ser ciertas para todos los casos, como en:

$$(\forall P)(P.costa\ \neq "Caribe")$$

Que significa, "para todas las tuplas de Provincias, la costa nunca es 'Caribe'".

Aquellas variables que están afectadas por un cuantificador se denominan **variables ligadas**. Las otras variables se denominan **variables libres**. Las únicas variables libres en una expresión de cálculo relacional deberían ser aquellas a la izquierda de la barra (|). Ejemplo:

$$\{P|Provincias(P) \wedge (\exists C)(Comunidades(C) \wedge (C.nombre = P.comunidad) \wedge C.capital = "Mérida"\}$$

P es la única variable libre, y será asociada a cada una de las tuplas de Provincias hasta encontrar todas aquellas que hacen cierta la expresión del lado derecho.

▪ Leyes de De Morgan

Las Leyes de De Morgan son leyes del cálculo proposicional que permiten la transformación de expresiones de uno a otro cuantificador.
Según estas leyes las siguientes expresiones son equivalentes:

$$(\exists x)(F(x)) \equiv \sim(\forall x)(\sim F(x))$$
$$(\forall x)(F(x)) \equiv \sim(\exists x)(\sim F(x))$$
$$(\exists x)(F_1(x) \wedge F_2(x)) \equiv \sim(\forall x)(\sim\big(F_1(x)\big) \wedge \sim(F_2(x)))$$
$$(\forall x)(F_1(x) \wedge F_2(x)) \equiv \sim(\exists x)(\sim\big(F_1(x)\big) \wedge \sim(F_2(x)))$$

▪ Expresiones y fórmulas

Las expresiones de en Cálculo Relacional de Tuplas deben seguir la siguiente estructura:

$$\{S_1 a_a, S_2 a_2, \ldots S_n a_n | F(S_1, S_2, \ldots S_m)\}\ m \geq n$$

Donde S_1, S_2, … S_n … Sm son variables de tupla; cada a_i es un atributo de la relación que define el dominio de S_i; y F es una fórmula. Las fórmulas bien formadas pueden ser atómicas o compuestas. Las fórmulas atómicas son cualquiera de las siguientes:

- R(S_i), donde S_i, es una variable de tupla y R una relación.

- $S_ia_1\ \theta\ S_ja_2$, donde S_i y S_j son variables de tupla; a_1 es un atributo de la relación que define a S_i, a_2 es un atributo de la relación que define a S_j y θ es un operador de comparación (=, ≠, >, ≥, <, ≤). Los atributos a_1 y a_2 deben poderse comparar con el operador θ.
- $S_ia_1\ \theta$ c, donde S_i es una variable de tupla; a_1 es un atributo de la relación que define a S_i, y θ es un operador de comparación.

Las fórmulas se definen recursivamente según las siguientes reglas:

- Una fórmula atómica es una fórmula
- Si F_1 y F_2 son fórmulas, también lo es su conjunción $F_1 \wedge F_2$, su disyunción $F_1 \vee F_2$, y su negación $\sim F_1$
- Si F es una fórmula que contiene la variable libre X, entonces (∃X)(F) y (∀X)(F) son también fórmulas.

- **Ejemplo de Cálculo Relacional de Tuplas**

a) Lista las provincias de Andalucía que tienen más de 1 000 000 habitantes.

$$\{P.nombre \mid Provincias(P) \wedge P.comunidadAutónoma = \text{"Andalucía"} \wedge P.habitantes = 1000000\}$$

b) Lista las comunidades autónomas que tienen alguna provincia con costa en el Atlántico

$$\{C.nombre \mid ComunidadAutonoma(C) \wedge (\exists P)(Provincias(P) \wedge P.costa = \text{"}Atlántico\text{"})\}$$

c) Lista las comunidades autónomas que no tienen mar

$$\{C.nombre \mid ComunidadAutonoma(C) \wedge (\nexists P)(Provincias(P) \wedge P.costa \neq null)\}$$

d) Lista las comunidades autónomas que tienen costa en el Mediteráneo y en el Atlántico

$$\{C|ComunidadAutonoma(C) \wedge (\exists P)(Provincias(P) \wedge P.comunidadAutonoma = C.nombre \wedge P.costa = \text{"}Atlántica\text{"}) \wedge (\exists Q)(Provincias(Q) \wedge Q.comunidadAutonoma = C.nombre \wedge Q.costa = \text{"}Mediterránea\text{"})\}$$

e) Lista las ciudades que son capital de una comunidad autónoma pero no son capital de provincia

$$\{C.capital\ |ComunidadAutonoma(C) \wedge (\nexists P)(Provincias(P) \wedge (P.capital = C.capital))\}$$

▪ Expresiones seguras y no seguras

Decimo que una expresión de cálculo relacional **no es segura**, cuando puede generar un conjunto infinito. Por ejemplo:

$$\{P| \sim Provincias(P)\}$$

Se trata de una expresión que devuelve cualquier tupla que no esté en la relación Provincias. Para evita casos como este, es necesario añadir una restricción que asegure que todas las tuplas de la expresión E pertenecen a un *dominio*. Esta expresión se escribe *dom(E).*

Una expresión será **segura** cuando todas las tuplas que devuelve pertenecen al dominio de la expresión.

La potencia expresiva del cálculo relacional de tuplas restringido a expresiones seguras es equivalente a la del álgebra relacional. Esto significa que para cada expresión segura de cálculo relacional de tuplas existe una expresión equivalente en álgebra relacional.

○ Cálculo Relacional de Dominios

Hemos visto como el cálculo relacional de tuplas se basa en variables que representan tuplas de diferentes relaciones. Por su parte, el cálculo relacional de dominios se basa en variables cuyos valores están en el ámbito del dominio de determinados atributos en lugar de tuplas completas de las relaciones. Una expresión genérica de cálculo relacional de dominios se escribiría como:

$$\{d_1, d_2, \dots, d_n | F(d_1, d_2, \dots, d_m)\} \qquad m \geq n$$

Donde d_1, d_2, …, d_n, …, dm son variables de dominio y $F(d_1, d_2, \dots d_m)$ es una fórmula compuesta de átomos según alguno de los siguientes tipos:

- $R(d_1, d_2, \dots, d_n)$ donde R es una relación de grado n y cada d es una variable de dominio.
- $d_i \theta d_j$, donde d_i y d_j son variables de dominio y θ, un operador de comparación. Los dominios de d_i y d_j deben poseer valores que puedan ser comparados con el operador θ.
- $d_i \theta c$, donde di es una variable de dominio, θ un operador de comparación, y c es una constante del dominio de d_i.

La composición de estos átomos para formar las fórmulas de las expresiones de cálculo relacional de dominios sigue las siguientes reglas recursivas:

- Un átomo es una fórmula
- Si F_1 y F_2 son fórmulas, también lo será la conjunción de $F_1 \wedge F_2$, su disyunción $F_1 \vee F_2$, y su negación ~ F_1.
- Si F es una fórmula que incluye la variable X, entonces (∃X)(F) y (∀X)(F) también son fórmulas.

Podemos halar de expresiones seguras y no seguras del mismo modo que lo hacíamos en cálculo relacional de tuplas. Además, el cálculo relacional de dominios restringido a expresiones seguras tiene una experiencia expresiva equivalente al cálculo relacional de tuplas restringido a expresiones seguras. Es decir, cualquier expresión segura de cálculo relacional de dominios tienen una fórmula segura equivalente en cálculo relacional de tuplas, y por tanto, también tendrá una expresión equivalente en álgebra relacional.

- Ejemplo de Cálculo Relacional de Dominios

f) Lista las provincias de Andalucía que tienen más de 1 000 000 habitantes.

$$\{nombre \mid (\exists Ca)(\exists Co)(\exists h)(Provincias(nombre, \text{"Andalucía"}, Ca, h, Co) \wedge h \geq 1000000\}$$

g) Lista las comunidades autónomas que tienen alguna provincia con costa en el Atlántico

$$\{nombre \mid (\exists p)(\exists ca)(\exists h)(\ Provincias(p, nombre, ca, h, \text{"}Atlántico\text{"}))\}$$

h) Lista las comunidades autónomas que no tienen mar

$$\{nombre \mid \sim(\exists p)\sim(\exists ca)\sim(\exists h)\sim(\exists costa)(Provincias(p, nombre, ca, h, costa) \wedge costa \neq null)\}$$

i) Lista las comunidades autónomas que tienen costa en el Mediterráneo y en el Atlántico

$$\{nombre \mid (\exists p)(\exists ca)(\exists h)(\ Provincias(p, nombre, ca, h, \text{"}Atlántico\text{"})) \\ \wedge (\exists p')(\exists ca')(\exists h')(\ Provincias(p', nombre, ca', h', \text{"}Mediterráneo\text{"}))\}$$

3 Otros lenguajes

El cálculo relacional es un lenguaje difícil de comprender y utilizar, pero la comunidad en torno a las bases de datos reconoció que su propiedad de ser no procedural es muy deseable en un lenguaje; por lo que comenzó una búsqueda de otras técnicas no procedurales más fáciles de usar. Así aparecieron otros dos tipos de lenguajes relacionales: los orientados a transformaciones y los lenguajes gráficos.

Los lenguajes orientados a transformaciones son un conjunto de lenguajes no procedurales que usa relaciones para transformar los datos de entrada en las salidas solicitadas. Estos lenguajes proporcionan estructuras fáciles de utilizar para indicar el resultado esperado en términos del conocimiento que se tiene de la base de datos. SQUARE, SEQUEL y su descendiente SQL pertenecen a este conjunto de lenguajes, siendo SQL el lenguaje de bases de datos relacionales más popular.

Los lenguaje gráficos proporcionan al usuario una representación gráfica del esquema de la relación. El usuario lo rellena con un ejemplo de lo que busca y el sistema devuelve los datos solicitados en ese formato. Query-By-Example (QBE) es un ejemplo de este tipo de lenguaje, de gran popularidad al ser el lenguaje de uso por defecto en la base de datos Access de la suite ofimática Microsoft Office.

Otra categoría son los **lenguajes de 4ª generación (4GL)**, que permiten construir aplicaciones a medida utilizando un conjunto limitado de comandos en un entorno amigable para el usuario, a menudo en un entorno basado en menús desplegables. Algunos sistemas aceptan una forma de lenguaje natural, una versión restringida del inglés. Se les denomina **lenguajes de 5ª generación (5GL)** si bien la investigación en este ámbito está aún en una etapa preliminar.

4 RESUMEN ESQUEMÁTICO

El álgebra relacional es un lenguaje procedimental de alto nivel, que se utiliza para construir nuevas relaciones a partir de las relaciones existentes en una base de datos. El cálculo relacional es un lenguaje no procedimental, que se utiliza para definir una relación mediante una fórmula con términos de otras relaciones existentes en una base de datos. Matemáticamente, el álgebra relacional y el cálculo relacional son lenguajes equivalentes, ya que para expresión en uno existe una expresión equivalente en el otro (y viceversa).

El cálculo relacional se utiliza como medida estándar de la capacidad expresiva de los lenguajes relacionales. Decimos que un lenguaje es relacionalmente completo cuando es capaz de producir cualquier relación que pueda ser definidia mediante el cálculo relacional. La mayoría de los lenguajes de consulta utilizados en los SGBD son relacionalmente completos, pero además son capaces de construir relaciones que van más allá de la capacidad del cálculo relacional o del álgebra relacional.

Las cinco operaciones básicas del álgebra relacional: Selección, Proyección, Producto Cartesiano, Unión y Diferencia de conjuntos son suficiente para construir la mayor parte de las operaciones de recuperación de datos que se pueden necesitar. Además, las operaciones derivadas: Join, Intersección y División permiten definir búsquedas de un modo más preciso.

El cálculo relacional es un lenguaje basado en predicados. Hay dos tipos de cálculo relacional: el cálculo relacional de tuplas, y el cálculo relacional de dominios.

En el cálculo relacional de tuplas se buscan aquellas tuplas para los que el predicado dado es cierto. Una variable de tupla es una variable cuyo dominio son las tuplas de determinada relación.

En el cálculo relacional de dominios se buscan variables que cumplan un predicado, donde el dominio de estas variables es el de los atributos de las relaciones involucradas, y no las tupas al completo como en el cálculo relacional de tuplas.

Es posible clasificar los lenguajes de manipulación de datos en procedurales o no procedurales, orientados a la transformación, gráficos, de cuarta generación y de quinta generación.

5 GLOSARIO

Cierre (Álgebra)

Un conjunto de elementos está cerrado con respecto a una o varias operaciones, si todos los resultados posibles de esas operaciones pertenecen al conjunto. El cierre de un conjunto con respecto a una o varias operaciones, es el conjunto de elementos que le faltan para considerarse cerrado. Por ejemplo: los números negativos son el cierre de los números naturales con respecto a la resta.

Compatibilidad para la unión

Dos relaciones son compatibles para la unión cuando tienen el mismo número de atributos, y el dominio de sus atributos coincide para cada pareja de atributos de una y otra relación tomados de manera ordenado. Esta propiedad se llama así porque es necesaria para que la operación de Unión sea cerrada con respecto al modelo relacional, pero también es necesaria en otras operaciones como la Intersección o la Diferencia de conjuntos.

Cuantificador existencial

El cuantificador existencial $\exists$, es una expresión matemática que, acompañado de una variable x, indica que el predicado al que se adjunta se cumple para al menos un valor de esa variable.

Cuantificador universal

El cuantificador universal $\forall$, es una expresión matemática que, acompañado de una variable x, indica que el predicado al que se adjunta se cumple para todos los valores de esa variable.

Dominio

Conjunto de valores que puede tomar una determinada variable. El dominio está relacionado con el concepto de tipo en los lenguajes de programación, pero generalmente es más específico. Por ejemplo, los valores “Cerdo” o “asdfghjk” encajan en el tipo *String*, pero no forman parte del dominio “nombres femeninos en español”.

Lenguajes

4GL (de cuarta generación)

Los lenguaje de cuarta generación incorporan elementos semánticos de modo que es posible definir acciones describiendo el resultado a obtener en lugar de la sucesión iterativa de pasos típica de los lenguajes procedimentales de tercera generación. Suelen estar especializados en un propósito específico.

5GL (de quinta generación)

Los lenguajes de quinta generación se basan en técnicas de inteligencia artificial que permiten la interpretación del lenguaje natural. Hoy día están en fase de desarrollo y experimentación.

Gráficos

Los lenguajes gráficos reemplazan las expresiones verbales y matemáticas de los lenguajes de programación tradicionales por diagramas o elementos visuales.

Orientados a transformaciones

Son un tipo de lenguajes no procedurales, de cuarta generación, que definen el resultado indicando las transformaciones que se deben realizar sobre los datos.

Operaciones

Agregadas

En álgebra relacional son operaciones avanzadlas que se realizan sobre los atributos de múltiples tuplas, y no sobre los de una única tuplas. Ejemplos de operaciones agregadas son CONTAR, MEDIA o DESVIACIÓN TÍPICA.

(de) Agrupamiento

Son operaciones de agrupamiento las que sirven para definir subconjuntos o grupos, en base a la posesión de valores comunes en los atributos indicados por el operador.

Básicas

Las operaciones básicas del álgebra relacional son aquellas que no pueden ser expresadas en base a otras operaciones.

Binarias

Son operaciones que producen una relación a partir de dos relaciones operando.

Derivadas

Son aquellas que pueden ser expresadas como una combinación de operaciones básicas.

Unarias

Las operaciones unarias producen una relación a partir de transformaciones realizadas en una única relación.

Predicado

Una expresión lógica que puede tomar diferentes valores de verdad en función de los valores asignados a las variables que forman parte de ella.

Proposición

Una proposición es un caso particular de predicado cuando se le asignan valores a las variables que aparecen en él. La proposición tendrá, por tanto, un valor de vedad determinado.

7 BIBLIOGRAFÍA BÁSICA

E.F. CODD (1970). A RELATIONAL MODEL OF DATA FOR LARGE SHARED DATA BANKS. COMMUNICATIONS OF THE ACM, 13(6), 377-387

E.F. CODD (1972). RELATIONAL COMPLETENESS OF DATA BASE SUBLANGUAGES. DATA BASE SYSTEMS,COURANT COMPUTER SCIENCE SYMPOSYUM 6TH (R. RUSTIN, ED.), ENGLEWOOD CLIFFS,NJ: PRENTICE HALL

T. CONNOLLY Y C.BEGG. DATABASE SYSTEMS: A PRACTICAL APPROACH TO DESIGN,IMPLEMENTATION, AND MANAGEMENT (6TH ED.). CAPÍTULO 5.