

Álgebra y Cálculos Relacionales

Mariano Beiró

Dpto. de Computación - Facultad de Ingeniería (UBA)

5 de abril de 2022

Temas

1 Introducción

2 Álgebra Relacional

- Operaciones básicas
- Operaciones adicionales: Junta externa
- Ejercicios

3 Cálculo Relacional

- Cálculo Relacional de Tuplas
- Cálculo Relacional de Dominios

4 Completitud Relacional

5 Bibliografía

1 Introducción

2 Álgebra Relacional

- Operaciones básicas
- Operaciones adicionales: Junta externa
- Ejercicios

3 Cálculo Relacional

- Cálculo Relacional de Tuplas
- Cálculo Relacional de Dominios

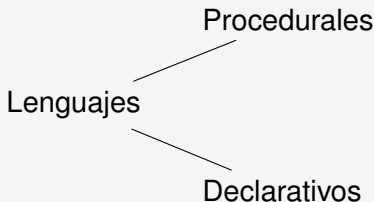
4 Completitud Relacional

5 Bibliografía

Introducción

Lenguajes de manipulación de datos (DML)

- Para interactuar con un modelo es necesario utilizar un **lenguaje**.
- Los lenguajes que permiten extraer información de un modelo de datos se denominan **lenguajes de manipulación de datos**, o DML (*Data Manipulation Languages*).



- Los lenguajes **procedurales** indican un procedimiento a seguir, utilizando **operaciones** que indican cómo manipular los datos.
- Los lenguajes **declarativos** indican qué resultado se quiere obtener, sin especificar cómo hacerlo.
- Los lenguajes procedurales se consideran de más bajo nivel.

Introducción

Lenguajes del modelo relacional

- En el modelo relacional existen distintos DML's, algunos de ellos formales y algunos prácticos.
- El lenguaje práctico más conocido es SQL (*Structured Query Language*), y es declarativo.
- Los lenguajes formales del modelo relacional son:
 - El **álgebra relacional** (procedural).
 - El **cálculo relacional** (declarativo).

1 Introducción

2 Álgebra Relacional

- Operaciones básicas
- Operaciones adicionales: Junta externa
- Ejercicios

3 Cálculo Relacional

- Cálculo Relacional de Tuplas
- Cálculo Relacional de Dominios

4 Completitud Relacional

5 Bibliografía

Álgebra Relacional

Características

pero relaciones y me devuelve una relación

- Es un lenguaje procedural.
- Propuesto por E. Codd en 1970, se lo considera parte integral del modelo relacional.
- Su utilidad radica en que:
 - Provee un marco formal de operaciones para el modelo relacional.
 - Se emplea como base para optimizar la ejecución de consultas.
- El álgebra relacional especifica los procedimientos de consulta de datos a partir de un conjunto de **operaciones**.
- Una operación -en el contexto del modelo relacional- es una función cuyos operandos son una o más relaciones, y cuyo resultado es también una relación.

$$O : R_1 \times R_2 \times \dots \times R_n \rightarrow S$$

- La **aridad** es la **cantidad de operandos** que toma una operación.
- Las operaciones del álgebra relacional pueden **combinarse** entre ellas para formar una **expresión**.

Álgebra Relacional

Recursos utilizados

- 2010 World Cup Dataset

<https://github.com/chinovieza/worldcup2010>

- RelaX - Relational algebra calculator

<http://dbis-uibk.github.io/relax/>



RelaX - relational algebra calculator

Calculator Feedback Help

RelaX - relational algebra calculator

calculates any relational algebra statement like $(\sigma_{a > 42}(A)) \bowtie (\pi_{a, b}(B))$ on a set of relations.

get started

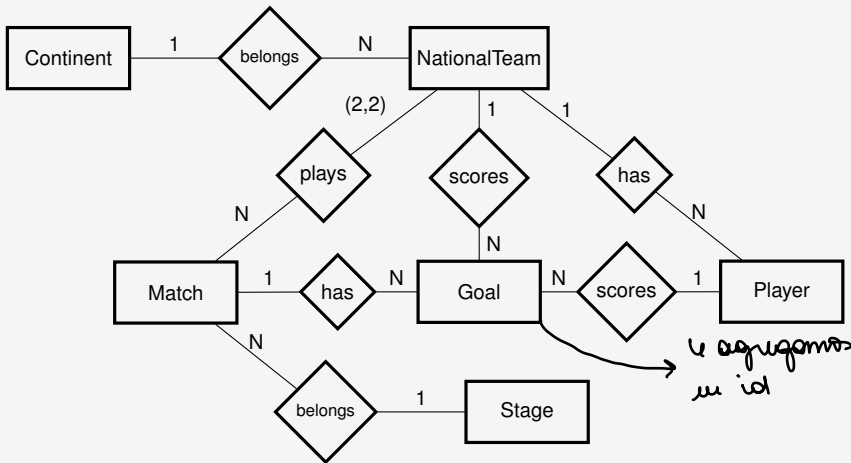


Link con dataset precargado: <https://dbis-uibk.github.io/relax/calc/gist/d617f36daeccd2135b124cd1a017429b>.

Álgebra Relacional

Recursos utilizados: 2010 World Cup Dataset

■ Modelo ER simplificado (sin atributos):



Álgebra Relacional

Recursos utilizados: 2010 World Cup Dataset

■ *Esquema de base de datos relacional:*

- Continent(id, name)
(1, 'Africa')
- NationalTeam(id, name, group, short_name, continent)
(1, 'South Africa', 'A', 'RSA', 0)
- Match(id, home, away, match_datetime_gmt, stage)
(1, 1, 2, '2010-06-11 14:00:00', 1)
- Player(id, name, birth_date, height, playing_position, local_club, national_team, national_team_tshirt)
(53, 'Edinson Cavani', '1987-02-14', 188, 'FW', 'Palermo [ITA]', 3, 7)
- Score(id, match_id, team_id, player_id, minute, score_type)
(1, 1, 1, 8, '55', 1) (Corresponde a la entidad Goal).
- Stage(id, name)
(3, 'Quarter-finals')
- Asumiremos que “name” es siempre clave candidata.
- Los tipos de score son: 1-normal; 2-penal; 3-tiempo adic.; 4-tiempo supl.; 5-gol en contra (se asigna al equipo contrario); 6-gol en serie de penales; 7-gol errado en serie de penales (no cuenta como gol).

1 Introducción

2 Álgebra Relacional

■ Operaciones básicas

■ Operaciones adicionales: Junta externa

■ Ejercicios

3 Cálculo Relacional

■ Cálculo Relacional de Tuplas

■ Cálculo Relacional de Dominios

4 Completitud Relacional

5 Bibliografía

Selección

↪ filtra por una condición

- El operador de **selección** (σ) es un operador unario. [ELM16 8.1.1]

$$\sigma_{cond} : R \rightarrow S$$

- Dada una relación $R(A_1, A_2, \dots, A_n)$ y una condición que se aplica a cada tupla de R , $\sigma_{cond}(R)$ selecciona aquellas tuplas de R para las cuales la condición es verdadera.

PELÍCULAS

nombre_película	año	nombre_director	cant_oscars
Kill Bill	2003	Quentin Tarantino	0
Django Unchained	2012	Quentin Tarantino	2
Star Wars III	2005	George Lucas	0
Coco	2017	Lee Unkrich	2

$$\downarrow \sigma_{cant_oscars \geq 1}(\text{Películas})$$

nombre_película	año	nombre_director	cant_oscars
Django Unchained	2012	Quentin Tarantino	2
Coco	2017	Lee Unkrich	2

Selección

Condiciones

- Utilizaremos condiciones atómicas de la forma:
 - $A_i \odot A_j$
 - $A_i \odot c$, con $c \in \text{dom}(A_i)$
- En donde \odot debe ser un operador de comparación:
 - $=, \neq$
 - $>, \geq, <, \leq$ (sólo para atributos cuyos dominios están ordenados)
- Una condición se construye combinando condiciones atómicas con los operadores lógicos **and** (\wedge), **or** (\vee) y **not** (\neg).

Ejemplo: World Cup 2010

Seleccionar aquellos jugadores del mundial que pertenecen al club "Barcelona [ESP]" y nacieron antes de 1990.

Respuesta

$\sigma_{(\text{local_club} = \text{"Barcelona [ESP]"} \wedge (\text{birth_date} < \text{"1990-01-01"}))}(\text{Player})$

Proyección

se queda con columnas [ELM16 8.1.2]

- El operador de **proyección** (π) es también un operador unario.

$$\pi_L : R \rightarrow S$$

los atributos que estoy eligiendo

- Dada una relación $R(A_1, A_2, \dots, A_n)$ y una lista de atributos $L = (L_1, L_2, \dots, L_k)$, con $L_i \in (A_1, A_2, \dots, A_n)$, $\pi_L(R)$ devuelve una relación cuyas tuplas representan los posibles valores de los atributos de L en R .

- Podemos pensar que lo que hace es *proyectar* cada tupla de R a un espacio de menor dimensión en que sólo se conservan los atributos que están en L .

PELÍCULAS

nombre_película	año	nombre_director
Kill Bill	2003	Quentin Tarantino
Django Unchained	2012	Quentin Tarantino
Star Wars III	2005	George Lucas
Coco	2017	Lee Unkrich

$\pi_{\text{nombre_director}}(\text{Películas})$

nombre_director
Quentin Tarantino
George Lucas
Lee Unkrich

Proyección

- El *orden* de los atributos en la relación resultado es el mismo orden en que figuran en L .
- El operador de proyección siempre *remueve tuplas duplicadas*, ya que su resultado debe ser también una relación válida.

Ejemplo: World Cup 2010

Liste las posiciones de juego de los jugadores.

Respuesta

$\pi_{\text{playing_position}}(\text{Player})$

la idea es ver cuál es el dominio del atributo

Secuencias de operaciones. Asignación (\leftarrow) [ELM16 8.1.3]

- Supongamos que queremos listar los nombres de los directores que han ganado algún Oscar.

PELÍCULAS

nombre_película	año	nombre_director	cant_oscars
Kill Bill	2003	Quentin Tarantino	0
Django Unchained	2012	Quentin Tarantino	2
Star Wars III	2005	George Lucas	0
Coco	2017	Lee Unkrich	2

$Temp \leftarrow \sigma_{cant_oscars > 0}(Películas)$

$Directores_Oscar \leftarrow \pi_{nombre_director}(Temp)$

- Podemos también hacerlo en un único paso:

$Directores_Oscar \leftarrow \pi_{nombre_director}(\sigma_{cant_oscars > 0}(Películas))$

DIRECTORES_OSCAR

nombre_director
Quentin Tarantino
Lee Unkrich

En Relax
usamos "="
en lugar de " \leftarrow "

Redenominación

[ELM16 8.1.3]

- El operador de **redenominación** (ρ) permite modificar los nombres de los atributos de una relación y/o el nombre de la relación misma.
- Nos permite preparar el resultado para la realización de una operación posterior.
- Dada una relación $R(A_1, A_2, \dots, A_n)$, un nuevo nombre de relación S y una lista de n nombres de atributo (B_1, B_2, \dots, B_n) , $\rho_{S(B_1, B_2, \dots, B_n)}(R)$ produce una relación de nombre S y atributos (B_1, B_2, \dots, B_n) cuyas tuplas coinciden con las tuplas de R .
- $\rho_S(R)$ sólo cambia el nombre de la relación R por S .

PELÍCULAS

nombre_película	cant_oscars
Kill Bill	0
Django Unchained	2
Star Wars III	0
Coco	2

 $\rho_{\text{Films}(\text{film_name}, \text{n_oscars})}(\text{Películas})$

FILMS

film_name	n_oscars
Kill Bill	0
Django Unchained	2
Star Wars III	0
Coco	2

Operaciones de conjuntos

Unión

[ELM16 8.2.1]

- Dadas dos relaciones $R(A_1, A_2, \dots, A_n)$ y $S(B_1, B_2, \dots, B_n)$, la **unión** $R \cup S$ es una **relación que contiene a todas las tuplas de R y de S** .
- Es necesario que R y S tengan el mismo grado.
- Además, para calcular $R \cup S$ las relaciones R y S deben coincidir en sus atributos en lo que respecta al dominio. Es decir, $dom(A_i) = dom(B_i)$. Esta condición se denomina **compatibilidad de unión** o **compatibilidad de tipo**.
- Por convención, en la relación resultado el listado de atributos coincide con el de R : (A_1, A_2, \dots, A_n) .

USUARIOS1

id1	nombre1
3	Juan
5	Martín
18	Marta

USUARIOS2

id2	nombre2
7	Marta
5	Martín

$$\xrightarrow{\text{Usuarios1} \cup \text{Usuarios2}}$$

id1	nombre1
3	Juan
5	Martín
18	Marta
7	Marta

Operaciones de conjuntos

Intersección

[ELM16 8.2.1]

- Dadas dos relaciones $R(A_1, A_2, \dots, A_n)$ y $S(B_1, B_2, \dots, B_n)$, la **intersección** $R \cap S$ conserva las tuplas que se encuentran presentes tanto en R como en S .
- R y S deben tener el mismo grado.
- Al igual que la unión, la intersección requiere compatibilidad de tipo.
- El listado de atributos de la relación resultado será (A_1, A_2, \dots, A_n) .

USUARIOS1

id1	nombre1
3	Juan
5	Martín
18	Marta

USUARIOS2

id2	nombre2
7	Marta
5	Martín

$$\xrightarrow{\text{Usuarios1} \cap \text{Usuarios2}}$$

id1	nombre1
5	Martín

Operaciones de conjuntos

Diferencia

[ELM16 8.2.1]

- Dadas dos relaciones $R(A_1, A_2, \dots, A_n)$ y $S(A_1, A_2, \dots, A_n)$, la **diferencia** $R - S$ conserva sólo aquellas tuplas de R que no pertenecen a S .
- R y S deben tener el mismo grado.
- También requiere compatibilidad de tipo.
- El listado de atributos de la relación resultado será (A_1, A_2, \dots, A_n) .

USUARIOS1

id1	nombre1
3	Juan
5	Martín
18	Marta

USUARIOS2

id2	nombre2
7	Marta
5	Martín

$$\xrightarrow{\text{Usuarios1} - \text{Usuarios2}}$$

id1	nombre1
3	Juan
18	Marta

- Las tres operaciones de conjuntos que definimos son operaciones binarias.

Producto cartesiano

[ELM16 8.2.2]

- Dadas dos relaciones $R(A_1, A_2, \dots, A_n)$ y $S(B_1, B_2, \dots, B_m)$, el **producto cartesiano** $R \times S$ produce una nueva relación T cuyas tuplas son *todas* aquellas de la forma $(t_1, t_2, \dots, t_n, t_{n+1}, t_{n+2}, \dots, t_{n+m})$, con $(t_1, t_2, \dots, t_n) \in R$ y $(t_{n+1}, t_{n+2}, \dots, t_{n+m}) \in S$.
- El esquema de la relación resultante T es $(A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_m)$. Salvo...
- ... si algún atributo A_i tiene el mismo nombre que un atributo B_j . En ese caso, la convención será que en el resultado los atributos se llamarán " $R.A_i$ " y " $S.B_j$ ". En el caso de estar calculando $R \times R$, llamaremos a los atributos $R1.A_i$ y $R2.A_i$.
- Aunque generalmente debe ser acompañada de alguna selección para reducir las combinaciones del resultado.
- El producto cartesiano no requiere compatibilidad de tipos.

Producto cartesiano

PELÍCULAS

nombre_película	nombre_director
Kill Bill	Quentin Tarantino
Django Unchained	Quentin Tarantino
Star Wars III	George Lucas
Coco	Lee Unkrich

ACTUACIONES

nombre_película	nombre_actor
Kill Bill	Uma Thurman
Star Wars III	Natalie Portman

↓ Películas × Actuaciones

Películas.nombre_película	nombre_director	Actuaciones.nombre_película	nombre_actor
Kill Bill	Quentin Tarantino	Kill Bill	Uma Thurman
Kill Bill	Quentin Tarantino	Star Wars III	Natalie Portman
Django Unchained	Quentin Tarantino	Kill Bill	Uma Thurman
Django Unchained	Quentin Tarantino	Star Wars III	Natalie Portman
Star Wars III	George Lucas	Kill Bill	Uma Thurman
Star Wars III	George Lucas	Star Wars III	Natalie Portman
Coco	Lee Unkrich	Kill Bill	Uma Thurman
Coco	Lee Unkrich	Star Wars III	Natalie Portman

Producto cartesiano

■ ¿Cómo hacemos, dadas las relaciones

■ Películas(nombre_película, nombre_director)

■ Actuaciones(nombre_película, nombre_actor)

para obtener las tuplas (nombre_película, nombre_director, nombre_actor) que representan la coparticipación de actores y directores en películas?

PELÍCULAS

nombre_película	nombre_director
Kill Bill	Quentin Tarantino
Django Unchained	Quentin Tarantino
Star Wars III	George Lucas
Coco	Lee Unkrich

ACTUACIONES

nombre_película	nombre_actor
Kill Bill	Uma Thurman
Star Wars III	Natalie Portman

↓ $\sigma_{\text{Películas.nombre_película}=\text{Actuaciones.nombre_película}}(\text{Películas} \times \text{Actuaciones})$

Películas.nombre_película	nombre_director	Actuaciones.nombre_película	nombre_actor
Kill Bill	Quentin Tarantino	Kill Bill	Uma Thurman
Star Wars III	George Lucas	Star Wars III	Natalie Portman

Producto cartesiano

Ejemplo: World Cup 2010

Liste los países con los que jugó la Selección Argentina.

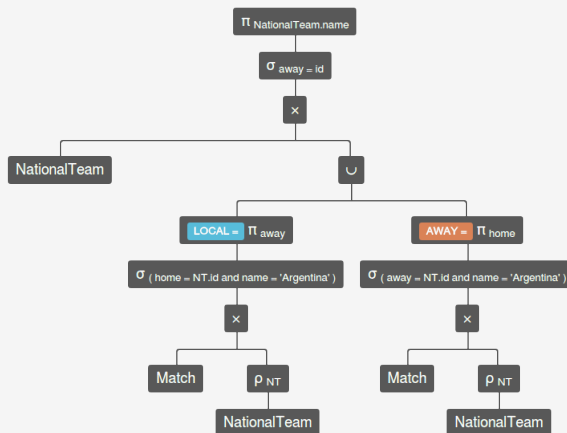
Respuesta

$$\begin{aligned}
 LOCAL &\leftarrow \pi_{away}(\sigma_{(home=NT.id) \wedge (name="Argentina")}(Match \times \rho_{NT}(NationalTeam))) \\
 AWAY &\leftarrow \pi_{home}(\sigma_{(away=NT.id) \wedge (name="Argentina")}(Match \times \rho_{NT}(NationalTeam))) \\
 \pi_{name}(\sigma_{away=id}(NationalTeam \times (LOCAL \cup AWAY)))
 \end{aligned}$$

Árboles de consulta

[ELM16 8.3.5]

- Para cada expresión del álgebra relacional se puede construir un **árbol de consulta** que representa el orden de ejecución.
- Para el ejemplo anterior sobre el producto cartesiano:



- La operación de **junta** combina un producto cartesiano con una selección. Dadas dos relaciones $R(A_1, A_2, \dots, A_n)$ y $S(B_1, B_2, \dots, B_m)$ y una condición, la junta $R \bowtie_{cond} S$ selecciona del producto cartesiano $R \times S$ las tuplas que cumplen la condición.
- No se admite cualquier tipo de condición de selección, sino sólo la conjunción de operaciones atómicas que incluyen columnas de ambas relaciones, es decir, de la forma:

- $A_i \odot B_j$

En donde \odot debe ser un operador de comparación:

- $=, \neq$

- $>, \geq, <, \leq$ (sólo para atributos cuyos dominios están ordenados)

Una condición se construye entonces combinando operaciones atómicas con el operador lógico **and** (\wedge).

Junta

- Ahora la combinación de Películas y Actuaciones se hace mucho más sencilla:

PELÍCULAS

nombre_película	nombre_director
Kill Bill	Quentin Tarantino
Django Unchained	Quentin Tarantino
Star Wars III	George Lucas
Coco	Lee Unkrich

ACTUACIONES

nombre_película	nombre_actor
Kill Bill	Uma Thurman
Star Wars III	Natalie Portman

↓ (Películas ⋈_{nombre_película=nombre_película} Actuaciones)

Películas.nombre_película	nombre_director	Actuaciones.nombre_película	nombre_actor
Kill Bill	Quentin Tarantino	Kill Bill	Uma Thurman
Star Wars III	George Lucas	Star Wars III	Natalie Portman

Junta

Ejemplo: World Cup 2010

Obtenga el listado de los nombres de los jugadores de la Selección Argentina.

Respuesta

$$PLAYER_TEAM \leftarrow NationalTeam \bowtie_{id=team_id} Player$$

$$\pi_{Player.name}(\sigma_{NationalTeam.name="Argentina"}(PLAYER_TEAM))$$

Junta

Tipos particulares de junta

[ELM16 8.3.2]

- El caso más general de operación de junta también se denomina **junta theta** (*theta join*).
- Cuando la junta sólo utiliza comparaciones de igualdad en sus condiciones atómicas, se denomina **junta por igual** (*equijoin*).
- En la junta por igual, el resultado dispondrá de pares de atributos distintos que poseerán información redundante. Para librarse de uno de ellos, se define la **junta natural**.

Junta

Junta Natural

- Para realizar una junta natural entre dos relaciones en reemplazo de una junta por igual, las mismas deben estar preparadas de manera que los pares de atributos (A_i, B_j) de cada condición atómica tengan el mismo nombre en una y otra relación. El resultado dispondrá de uno sólo de los atributos, conservando su nombre.
- La junta natural entre dos relaciones R y S se simboliza $R * S$.
- **¡Atención!** En la junta natural no se especifican las condiciones, por lo tanto todo par de atributos de igual nombre en una y otra relación será comparado por igual en la condición de selección implícita.
- Los atributos comparados en una junta se denominan **atributos de junta**.

Junta

Junta Natural

- Volviendo al caso de las Películas y Actuaciones:
 - Películas(nombre_película, nombre_director)
 - Actuaciones(nombre_película, nombre_actor)
- Las relaciones ya están preparadas para una junta natural.

PELÍCULAS

nombre_película	nombre_director
Kill Bill	Quentin Tarantino
Django Unchained	Quentin Tarantino
Star Wars III	George Lucas
Coco	Lee Unkrich

ACTUACIONES

nombre_película	nombre_actor
Kill Bill	Uma Thurman
Star Wars III	Natalie Portman

↓ (Películas * Actuaciones)

nombre_película	nombre_director	nombre_actor
Kill Bill	Quentin Tarantino	Uma Thurman
Star Wars III	George Lucas	Natalie Portman

Junta

Junta Natural: Ejemplo

Ejemplo: RENAPER

Personas(DNI, nombre, género, fecha_nacimiento)

HijoDe(DNI_padre, DNI_hijo)

CasadaCon(DNI1, DNI2, fecha_matrimonio)

Liste a todos los hijos de “Abraham Simpson” (suponga que no hay dos personas con ese nombre).

Respuesta

$PADRE \leftarrow \rho_{DNI_padre}(\pi_{DNI}(\sigma_{nombre="Abraham\ Simpson"}(Personas)))$

$HIJOS \leftarrow \rho_{DNI_hijo, nombre}(\pi_{DNI, nombre}(Personas))$

$\pi_{DNI_hijo, nombre}(PADRE * HijoDe * HIJOS)$

División

- Esta vez, primero el ejemplo...
- Nos interesa saber qué alumnos aprobaron los 3 TPs.

NOTAS

alumno	TP	nota
Pedro	1	7
Pedro	3	2
Juan	1	3
Juan	2	6
Juan	3	8
Walter	1	4
Walter	2	9
Walter	3	8

→

APROBADOS

alumno	TP
Pedro	1
Juan	2
Juan	3
Walter	1
Walter	2
Walter	3

REQUISITOS

TP
1
2
3

↓ (Aprobados ÷ Requisitos)

alumno
Walter

División

[ELM16 8.3.4]

- Es una operación inversa al producto cartesiano.
- Partimos de una relación $R(A_1, A_2, \dots, A_n)$ y una relación $S(B_1, B_2, \dots, B_m)$ cuyos atributos están incluidos en los de R .
- Llamaremos $A = \{A_1, A_2, \dots, A_n\}$ y $B = \{B_1, B_2, \dots, B_m\}$. Entonces $B \subset A$.
- Llamaremos $Y = A - B$.
- Se define entonces la **división** $R \div S$ como la relación $T(Y)$ que contiene todas las tuplas t que cumplen que:
 - 1 t pertenece a $\pi_Y(R)$.
 - 2 Para cada tupla $t_S \in S$ existe una tupla $t_R \in R$ tal que $t_R[Y] = t$ y $t_R[B] = t_S$.
- Propiedad: T es la relación de mayor cardinalidad posible contenida en $\pi_Y(R)$ y que cumple que $T * S \subset R$.

División

Ejemplo: Tenistas

Tenistas(nombre_tenista, país, altura, diestro)

(‘Novak Djokovic’, ‘Serbia’, 1.88, True)

Torneos(nombre_torneo, tipo_torneo)

(‘Abierto de Australia’, ‘Grand Slam’)

Campeones(nombre_tenista, nombre_torneo, modalidad, año)

(‘Juan Martín del Potro’, ‘Torneo de Estocolmo’, ‘Single’, 2016)

Liste a aquellos tenistas que hayan ganado todos los torneos de tipo “Grand Slam” existentes al menos una vez.

Respuesta

$TORNEOS_GRAND_SLAM \leftarrow \pi_{nombre_torneo}(\sigma_{tipo_torneo = "Grand Slam"}(Torneos))$

$\pi_{nombre_tenista, nombre_torneo}(Campeones) \div TORNEOS_GRAND_SLAM$

Conjuntos completos de operadores

[ELM16 8.3.3]

- Hemos definido una serie de operadores básicos del álgebra relacional: $\sigma, \pi, \rho, \cup, \cap, -, \times, \bowtie, *, \div$.
- Sin embargo, existen subconjuntos de ellos que tienen la misma capacidad de expresión que todo el conjunto.
- A dichos subconjuntos se los denomina **conjuntos completos de operadores**.
- $\{\sigma, \pi, \rho, \cup, -, \times\}$ forman un conjunto completo de operadores.

¿Cómo se demuestra?

Mostrando que cada uno de los operadores restantes puede construirse a partir de estos seis.

1 Introducción

2 Álgebra Relacional

- Operaciones básicas
- **Operaciones adicionales: Junta externa**
- Ejercicios

3 Cálculo Relacional

- Cálculo Relacional de Tuplas
- Cálculo Relacional de Dominios

4 Completitud Relacional

5 Bibliografía

Operaciones adicionales

[ELM16 8.4]

- Existen operaciones frecuentes de bases de datos que no pueden ser expresadas en el álgebra relacional básica.
- Se han propuesto numerosos operadores para extender el álgebra relacional, entre ellos:
 - La proyección generalizada.
 - La agregación.
 - La **junta externa**.
- Sólo presentaremos aquí la *junta externa*.

Junta externa

- Volvamos al ejemplo de las películas y los actores.

PELÍCULAS

nombre_película	nombre_director
Kill Bill	Quentin Tarantino
Django Unchained	Quentin Tarantino
Star Wars III	George Lucas
Coco	Lee Unkrich

ACTUACIONES

nombre_película	nombre_actor
Kill Bill	Uma Thurman
Star Wars III	Natalie Portman

↓ (Películas $\bowtie_{\text{nombre_película}=\text{nombre_película}}$ Actuaciones)

Películas.nombre_película	nombre_director	Actuaciones.nombre_película	nombre_actor
Kill Bill	Quentin Tarantino	Kill Bill	Uma Thurman
Star Wars III	George Lucas	Star Wars III	Natalie Portman

- El resultado muestra las combinaciones de director y actor que trabajaron juntos, pero descarta las tuplas de la relación izquierda con las cuales no se combina ninguna tupla de la derecha.

Junta externa

[ELM16 8.4.4; SILB19 4.1.3]

- La **junta externa** evita que eso suceda, asegurando que las tuplas de una o ambas relaciones estén presentes en el resultado, aún cuando no puedan combinarse con ninguna tupla de la otra.
- Existen 3 tipos de junta externa:
 - Junta externa izquierda ($R \bowtie S$)
 - Junta externa derecha ($R \ltimes S$)
 - Junta externa completa ($R \ltimes S$)
- Dadas dos relaciones $R(A_1, A_2, \dots, A_n)$ y $S(B_1, B_2, \dots, B_m)$ y una condición, la junta externa $R[\bowtie, \ltimes, \ltimes]_{cond} S$ selecciona del producto $R \times S$ las tuplas que cumplen la condición, y añade...
 - ...una tupla $(t[A_1], t[A_2], \dots, t[A_n], NULL, NULL, \dots, NULL)$ de dimensión $n + m$ por cada tupla de $t \in R$ que no se encuentra en la proyección sobre (A_1, A_2, \dots, A_n) (**Junta externa izquierda, \bowtie**).
 - ...una tupla $(NULL, NULL, \dots, NULL, t[B_1], t[B_2], \dots, t[B_m])$ de dimensión $n + m$ por cada tupla de $t \in S$ que no se encuentra en la proyección sobre (B_1, B_2, \dots, B_m) (**Junta externa derecha, \ltimes**).
 - ...ambos tipos de tuplas descriptos (**Junta externa completa, \ltimes**).

1 Introducción

2 Álgebra Relacional

- Operaciones básicas
- Operaciones adicionales: Junta externa
- Ejercicios

3 Cálculo Relacional

- Cálculo Relacional de Tuplas
- Cálculo Relacional de Dominios

4 Completitud Relacional

5 Bibliografía

Ejercicios: World Cup 2010

Ejercicio 1

Liste el nombre de los continentes que no fueron representados por ningún equipo en los cuartos de final del Mundial.

Ejercicio 2

Liste el nombre de los jugadores que marcaron al menos 3 goles durante el Mundial.

Ejercicio 3

Liste el nombre y selección nacional de el/los jugadores más altos del Mundial.

1 Introducción

2 Álgebra Relacional

- Operaciones básicas
- Operaciones adicionales: Junta externa
- Ejercicios

3 Cálculo Relacional

- Cálculo Relacional de Tuplas
- Cálculo Relacional de Dominios

4 Completitud Relacional

5 Bibliografía

Cálculo Relacional

Características

- Es un lenguaje declarativo, de más alto nivel que el álgebra relacional.
- Al no ser procedural, no especifica un orden de operaciones a realizar.
- Está basado en la **lógica de predicados**.
- Presenta dos variantes:
 - El **cálculo relacional de tuplas**.
 - El **cálculo relacional de dominios**.
- El lenguaje SQL está inspirado en el cálculo relacional de tuplas.

Cálculo Relacional

Proposiciones, predicados y relaciones

■ Proposiciones:

- Rafael Nadal ganó el torneo de Roland Garros en 2009.
- Gabriela Sabatini ganó el Abierto de Estados Unidos en 1990.
- Andy Murray ganó el Campeonato de Wimbledon en 2016.
- Un conjunto de proposiciones que tienen la misma estructura puede tipificarse a través de un predicado.
 - [tenista] ganó [torneo] en [año]
- Un predicado es una función cuyo resultado es un valor de verdad: Verdadero (V) ó Falso (F)
 - $TenistaCampeón(nombre_tenista, nombre_torneo, año)$
 - Entonces:
 - $TenistaCampeón(Rafael Nadal, Roland Garros, 2011) = V$
 - $TenistaCampeón(Juan Martín del Potro, Roland Garros, 2011) = F$
- En el cálculo relacional, los esquemas de relación pueden pensarse como predicados.
 - Las bases de datos sólo almacenan proposiciones verdaderas.
 - Ergo, cada tupla t de una relación R predica que $R(t) = V$.

Cálculo Relacional

Lógica de predicados de primer orden

- La lógica de predicados de primer orden se basa en:
 - **Predicados**: Son funciones de una o más variables cuyo resultado es un valor de verdad (V ó F).
 - Ejemplo: $p(m, n)$.
 - **Operaciones** entre predicados.
 - $\wedge, \vee, \neg, \rightarrow$
 - Combinando los predicados con operaciones se obtienen predicados más complejos
 - Ejemplo: $(p(m, n) \wedge \neg q(m)) \vee q(n)$.
 - **Quantificadores** de variables.
 - Cuantificador universal: $(\forall m)q(m)$. Es verdadero si para cualquier valor de m el predicado $q(m)$ es verdadero.
 - Cuantificador existencial: $(\exists m)q(m)$. Es verdadero si existe al menos un valor de m para el cual el predicado $q(m)$ es verdadero.

1 Introducción

2 Álgebra Relacional

- Operaciones básicas
- Operaciones adicionales: Junta externa
- Ejercicios

3 Cálculo Relacional

- Cálculo Relacional de Tuplas
- Cálculo Relacional de Dominios

4 Completitud Relacional

5 Bibliografía

Cálculo Relacional de Tuplas

Predicados y operaciones

[ELM16 8.6; SILB19 27.1; CONN15 5.2.1]

- En el **cálculo relacional de tuplas** las variables representan tuplas.
- Un **predicado simple** es una función de una tupla o de atributos de tuplas, cuyo resultado es un valor de verdad (V ó F). Se admiten como predicados simples:
 - $R(t)$, en donde R es una relación
 - $t_1.A_i \odot t_2.A_j$
 - $t.A_i \odot c$, con $c \in \text{dom}(A_i)$
 - En donde \odot debe ser un operador de comparación:
 - $=, \neq$
 - $>, \geq, <, \leq$ (sólo para atributos cuyos dominios están ordenados)
- Las **operaciones** entre predicados admitidas son \wedge, \vee, \neg .

Una expresión del cálculo relacional de tuplas tiene la forma:
 $\{t_1.A_{i1}, t_1.A_{i2}, \dots, t_1.A_{i_{k_1}}, \dots, t_n.A_{i_{k_n}} | p(t_1, t_2, \dots, t_n, t_{n+1}, \dots, t_{n+m})\}$,
 en donde p es un predicado válido. $\{t_1, t_2, \dots, t_n\}$ deben ser *variables libres*, y $\{t_{n+1}, t_{n+2}, \dots, t_{n+m}\}$ deben ser *variables ligadas*.

Cálculo Relacional de Tuplas

Ejemplos

Ejemplo: World Cup 2010

Liste los nombres de los países que jugaron el Mundial 2010.

Respuesta

$\{n.name | NationalTeam(n)\}$

Ejemplo: World Cup 2010

Liste los nombres de los jugadores nacidos antes de 1980.

Respuesta

$\{p.name | Player(p) \wedge p.date_birth < 1980-01-01\}$

Cálculo Relacional de Tuplas

Cuantificadores

[ELM16 8.6.3; ELM16 8.6.6]

- Pero, ¿cómo hacemos si queremos listar a los jugadores que hicieron *algún* gol durante el mundial?
- Necesitamos de los **cuantificadores**.
 - **Cuantificador universal:** $(\forall t)p(t)$. Es verdadero si para cualquier tupla t el predicado $p(t)$ es verdadero.
 - **Cuantificador existencial:** $(\exists t)p(t)$. Es verdadero si existe al menos una tupla t para la cual el predicado $p(t)$ es verdadero.
- El listado de los nombres de los jugadores que hicieron goles se obtiene como:

Respuesta

$$\{p.name \mid Player(p) \wedge (\exists s)(Score(s) \wedge s.player_id = p.id)\}$$

Cálculo Relacional de Tuplas

Cuantificadores

- **Atención!** Una variable que fue cuantificada no puede aparecer seleccionada en el lado izquierdo de la barra ($|$), y toda variable que aparece sólo en el lado derecho debe estar cuantificada.
 - Las variables que fueron cuantificadas son **variables ligadas**.
 - Las variables que no fueron cuantificadas son **variables libres**.
- Reiteramos:

Una expresión del cálculo relacional de tuplas tiene la forma:
 $\{t_1.A_{11}, t_1.A_{12}, \dots, t_1.A_{1k_1}, \dots, t_n.A_{nk_n} | p(t_1, t_2, \dots, t_n, t_{n+1}, \dots, t_{n+m})\}$,
en donde p es un predicado válido. $\{t_1, t_2, \dots, t_n\}$ deben ser *variables libres*, y $\{t_{n+1}, t_{n+2}, \dots, t_{n+m}\}$ deben ser *variables ligadas*.

Cálculo Relacional de Tuplas

Cuantificadores

Ejemplo: World Cup 2010

Liste los nombres de los jugadores de la Selección Española.

Respuesta

$$\{p.name \mid Player(p) \wedge (\exists n)(NationalTeam(n) \wedge n.id = p.national_team \wedge n.name = "Spain")\}$$

Cálculo Relacional de Tuplas

Cuantificadores

Ejemplo: World Cup 2010

Liste el nombre del jugador más anciano del Mundial.

Respuesta

$$\{p.name \mid Player(p) \wedge (\forall \theta)(\neg Player(\theta) \vee \theta.birth_date \geq p.birth_date)\}$$

- Observemos que el cuantificador $\forall \theta$ necesita típicamente de una negación dentro de su expresión, para restringir el universo de θ 's sobre los que requerimos que la expresión sea verdadera. De lo contrario, el resultado estará vacío.

Cálculo Relacional de Tuplas

Expresiones seguras

[ELM16 8.6.8; SILB19 27.2.3]

- No toda expresión válida del cálculo de tuplas es una **expresión segura** (*safe expression*).
- Por ejemplo, la expresión...

$$\{p.name | \neg Player(p)\}$$

- ... no es una expresión segura. Producirá una cantidad infinita de tuplas con valores como “safsq” o 57.
- Una **expresión segura** es aquella que garantiza formalmente que producirá una cantidad finita de tuplas.
- Puede probarse que ésto es equivalente a garantizar que los valores de los atributos del resultado son parte del dominio de la expresión.

Cálculo Relacional de Tuplas

Expresiones seguras

Ejemplos:

- $\{p_1.nombre | (\exists p_2)(Persona(p_2) \wedge p_2.edad = p_1.edad)\}$

- ✗ Expresión no segura

- Probablemente queríamos

$$\{p_1.nombre | Persona(p_1) \wedge (\exists p_2)(Persona(p_2) \wedge p_2.edad = p_1.edad)\}$$

- $\{p_1.nombre | Empleado(p_1) \wedge (\nexists p_2)(Empleado(p_2) \wedge p_2.sueldo > p_1.sueldo)\}$

- ✓ Expresión segura

- $\{t.nombre | \neg((Cliente(t) \wedge Proveedor(t))\}$

- ✗ Expresión no segura

- Probablemente queríamos

$$\{t.nombre | (Cliente(t) \vee Proveedor(t)) \wedge \neg(Cliente(t) \wedge Proveedor(t))\}$$

- **Recomendación:** Cuidado cuando usamos cuantificadores ó negamos predicados!

Cálculo Relacional de Tuplas

Ejercicio: Tenistas

Tenistas(nombre_tenista, país, altura, diestro)

(‘Novak Djokovic’, ‘Serbia’, 1.88, True)

Torneos(nombre_torneo, tipo_torneo)

(‘Abierto de Australia’, ‘Grand Slam’)

Campeones(nombre_tenista, nombre_torneo, modalidad, año)

(‘Juan Martín del Potro’, ‘Torneo de Estocolmo’, ‘Single’, 2016)

Liste los nombres de los tenistas que ganaron todos los torneos de Grand Slam.

Respuesta

$$\{c.nombre_tenista \mid \text{Campeones}(c) \wedge$$

$$(\forall t)(\neg \text{Torneos}(t) \vee t.tipo_torneo \neq \text{"Grand Slam"}) \vee$$

$$(\exists c_2)(\text{Campeones}(c_2) \wedge$$

$$c_2.nombre_tenista = c.nombre_tenista \wedge$$

$$c_2.nombre_torneo = t.nombre_torneo))\}$$

1 Introducción

2 Álgebra Relacional

- Operaciones básicas
- Operaciones adicionales: Junta externa
- Ejercicios

3 Cálculo Relacional

- Cálculo Relacional de Tuplas
- Cálculo Relacional de Dominios

4 Completitud Relacional

5 Bibliografía

Cálculo Relacional de Dominios

Predicados, operaciones y cuantificadores

[ELM16 8.7]

- En el **cálculo relacional de dominios** las variables representan dominios, es decir que hacen referencia a los atributos.
- Un **predicado simple** es una función de un conjunto de dominios, cuyo resultado es un valor de verdad (V ó F). Se admiten como predicados simples:
 - $R(x_1, x_2, \dots, x_n)$, en donde $R(A_1, A_2, \dots, A_n)$ es una relación
 - $x_i \odot x_j$
 - $x_i \odot c$, con $c \in \text{dom}(A_i)$
 - En donde \odot debe ser un operador de comparación:
 - $=, \neq$
 - $>, \geq, <, \leq$ (sólo para atributos cuyos dominios están ordenados)
- Las **operaciones** entre predicados admitidas son \wedge, \vee, \neg .
- Se utilizan los **cuantificadores** con las mismas reglas que en el CRT.

Cálculo Relacional de Dominios

Predicados y operaciones

Una expresión del cálculo relacional de dominios tiene la forma:
 $\{x_1, x_2, \dots, x_n | p(x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_{n+m})\}$,
en donde p es un predicado válido. $\{x_1, x_2, \dots, x_n\}$ deben ser *variables libres*, y $\{x_{n+1}, x_{n+2}, \dots, x_{n+m}\}$ deben ser *variables ligadas*.

Cálculo Relacional de Dominios

Ejemplos

Ejemplo: World Cup 2010

Liste los nombres de los países que jugaron el Mundial 2010.

Respuesta

$$\{n | (\exists i)(\exists g)(\exists s)(\exists c)(NationalTeam(i, n, g, s, c))\}$$

Ejemplo: World Cup 2010

Liste los nombres de los jugadores nacidos antes de 1980.

Respuesta

$$\{n | (\exists i)(\exists b)(\exists h)(\exists p)(\exists l)(\exists t)(\exists s)(Player(i, n, b, h, p, l, t, s) \wedge b < 1980-01-01)\}$$

Cálculo Relacional de Dominios

Ejemplos

Ejemplo: RENAPER

Personas(DNI, nombre, género, fecha_nacimiento)

HijoDe(DNI_padre, DNI_hijo)

CasadaCon(DNI1, DNI2, fecha_matrimonio)

Liste a todos los hijos de “Abraham Simpson” (suponga que no hay dos personas con ese nombre).

Respuesta

$$\{ h, n_1 | (\exists d_1)(\exists g_1)(\exists f_1)(\exists d_2)(\exists n_2)(\exists g_2)(\exists f_2)(\exists p) \\ (\text{Personas}(d_1, n_1, g_1, f_1) \wedge \text{Personas}(d_2, n_2, g_2, f_2) \\ \wedge \text{HijoDe}(p, h) \wedge n_2 = \text{“Abraham Simpson”} \wedge h = d_1 \wedge p = d_2) \}$$

1 Introducción

2 Álgebra Relacional

- Operaciones básicas
- Operaciones adicionales: Junta externa
- Ejercicios

3 Cálculo Relacional

- Cálculo Relacional de Tuplas
- Cálculo Relacional de Dominios

4 Completitud Relacional

5 Bibliografía

Completitud Relacional

- E. Codd demostró la **equivalencia** entre el álgebra relacional básica y el cálculo relacional¹.
- Esta equivalencia implica que ambos lenguajes tienen el mismo **poder expresivo**.
 - Toda consulta expresable a través del cálculo relacional es también expresable en el álgebra relacional básica y viceversa.

¿Cómo se demuestra esta equivalencia?

Mostrando que cada uno de los operadores del álgebra relacional básica es expresable a través del cálculo relacional, y que una expresión genérica segura del cálculo relacional es expresable utilizando los operadores del álgebra relacional básica.

- A su vez, se dice que un lenguaje es **relacionalmente completo** cuando tiene la misma capacidad expresiva que el cálculo relacional.
 - El álgebra relacional básica es relacionalmente completa.

¹Restringido a expresiones seguras.

1 Introducción

2 Álgebra Relacional

- Operaciones básicas
- Operaciones adicionales: Junta externa
- Ejercicios

3 Cálculo Relacional

- Cálculo Relacional de Tuplas
- Cálculo Relacional de Dominios

4 Completitud Relacional

5 Bibliografía

Bibliografía

[ELM16] Fundamentals of Database Systems, 7th Edition.

R. Elmasri, S. Navathe, 2016.

Capítulo 8

[SILB19] Database System Concepts, 7th Edition.

A. Silberschatz, H. Korth, S. Sudarshan, 2019.

Capítulo 2.5 y 2.6, Capítulo 27 (online)

[CONN15] Database Systems, a Practical Approach to Design, Implementation and Management, 6th Edition.

T. Connolly, C. Begg, 2015.

Capítulo 5