

Cálculo Relacional

Base de Datos

Mónica Caniupán
mcaniupan@ubiobio.cl

Universidad del Bío-Bío

2020

Cálculo Relacional

- El cálculo relacional, a diferencia del AR, es un **lenguaje declarativo**
 - El CR nos permite expresar las consultas sin decir como obtener las respuestas
- Existen dos variantes del CR:
 - 1 Cálculo Relacional de Tuplas (CRT)
 - 2 Cálculo Relacional de Dominios (CRD)

Cálculo Relacional de Tuplas

- Una variable **tupla** es una variable que adopta como valores las tuplas de una relación
- Es decir, cualquier asignación de valores a una variable tupla tiene el mismo número y tipo de atributos
- **Ejemplo:** Consideremos la relación ALUMNOS:

ALUMNOS		
ID	NOMBRE	EDAD
11	maria	20

- Si X es una variable de tipo tupla entonces le podemos asignar la tupla en ALUMNOS:

$$X.ID = 11, X.NOMBRE = 'maria', X.EDAD = 20$$

Cálculo Relacional de Tuplas

- Una consulta en CRT es de la forma:

$$\{T \mid p(T)\}$$

donde T es una variable tupla y $p(T)$ denota una fórmula que describe T

- El resultado de esta consulta es el conjunto de todas las tuplas t para las cuales la fórmula $p(T)$ se evalúa como verdadera con $T = t$

Ejemplo: CRT

- Dada la relación ALUMNOS:

ALUMNOS		
ID	NOMBRE	EDAD
11	<i>pedro</i>	21
12	<i>luis</i>	22
13	<i>juan</i>	20

- La consulta: “Encontrar los alumnos con edad mayor a 20 años” se expresa por:

$$\{x \mid x \in ALUMNOS \wedge x.EDAD > 20\}$$

- La respuesta a la consulta es:

ID	NOMBRE	EDAD
11	<i>pedro</i>	21
12	<i>luis</i>	22

Sintaxis de Consultas en CRT

- Sean *Rel* el nombre de una relación, *R*, *S* variables tuplas con atributos *a*, *b* respectivamente, i.e. *R.a*, *S.b*
- Y sea *op* uno de los siguientes operadores: $\{<, >, =, \leq, \geq, \neq\}$
- Una **fórmula atómica** es:
 - $R \in Rel, \quad x \in ALUMNOS$
 - $R.a \text{ op } S.b, \quad x.EDAD = y.EDAD$
 - $R.a \text{ op } c, \text{ con } c \text{ constante}, \quad x.EDAD > 20$
 - $c \text{ op } R.a, \text{ con } c \text{ constante}, \quad 20 < x.EDAD$

Sintaxis de Consultas en CRT

- Una **fórmula** se define recursivamente de la siguiente manera:
 - Toda fórmula atómica p es una fórmula
 - Si p y q son fórmulas, también lo son:
 - $\neg p$,
 - $(p \wedge q)$,
 - $(p \vee q)$,
 - $(p \rightarrow q)$
 - $\exists R(p(R))$, donde R es una variable tupla
 - $\forall R(p(R))$, donde R es una variable tupla
- Los cuantificadores \exists, \forall limitan la variable R
- Una variable es **libre** en una fórmula (o subfórmula) si la (sub)fórmula no contiene ninguna ocurrencia de cuantificadores que la limiten
- En una consulta en CRT de la forma: $\{T \mid p(T)\}$, T es la única variable libre

Semántica de Consultas en CRT

- La semántica de las consultas nos permite saber cual es el conjunto respuesta para una consulta de la forma $\{T \mid p(T)\}$
- La respuesta a una consulta expresada en CRT es el conjunto de todas las tuplas t para las cuales la fórmula $p(T)$ es verdadera cuando a la variable T se le asigna el valor de t
- Aquí surgen algunas preguntas:
 - ¿Qué significa asignar valores a una variable?
 - ¿Cuáles son los posibles valores a asignar?

Semántica de Consultas en CRT

- Cada consulta se evalúa sobre una instancia de la BD
- Supongamos que cada variable libre de la fórmula $F = p(T)$ está ligada a un valor tupla
- Para una asignación dada de tuplas a variables con respecto a la instancia de BD, F se evalúa como verdadera si se cumple alguna de las siguientes condiciones:
 - F es la fórmula atómica $R \in Rel$ y a R se le asigna una tupla de la relación R
 - F es una comparación $R.a \text{ op } S.b$, $R.a \text{ op } c$, $c \text{ op } R.a$, con c constante y las tuplas asignadas a R e S tienen los valores de $R.a$ y $S.b$ que hacen que la comparación sea verdadera

Semántica de Consultas en CRT

- F es de la forma:
 - $\neg p$ y p no es verdadera,
 - $p \wedge q$ y ambas p y q son verdaderas,
 - $p \vee q$ y al menos una de ellas es verdadera,
 - $p \rightarrow q$ y q es verdadera siempre que p es verdadera
- F es de la forma $\exists R(p(R))$ y hay alguna asignación de tuplas a las variables libres de $p(R)$, incluyendo la variable R , que hace que la fórmula sea verdadera
- F es de la forma $\forall R(p(R))$ y existe alguna asignación de tuplas a las variables libres de $p(R)$ que hace que la fórmula $p(R)$ sea verdadera independiente de la tupla que se asigne a R

Ejemplo: Consultas en CRT

- Consideremos el siguiente esquema e instancia de BD:

CUR	
IDC	NOMBREC
1	<i>BD1</i>
2	<i>BD2</i>

INS		
ID	IDC	NOTA
10	1	70
10	2	85
11	2	80

ALUMNOS			
ID	NOMBRE	EDAD	CIUDAD
10	<i>luis</i>	20	<i>concepcion</i>
11	<i>pedro</i>	21	<i>chillan</i>
12	<i>antonio</i>	23	<i>concepcion</i>

Ejemplo: Consultas en CRT

- Esquema:

- $ALUMNOS(ID, NOMBRE, EDAD, CIUDAD)$
- $CUR(IDC, NOMBREC)$
- $INS(ID, IDC, NOTA)$

- Q_1 : Encontrar el nombre y la edad de los alumnos que viven en Concepción

$$\{x \mid \exists y \in ALUMNOS(y.CIUDAD = 'concepcion' \wedge x.NOMBRE = y.NOMBRE \wedge x.EDAD = y.EDAD)\}$$

- x es una variable tupla con dos atributos $NOMBRE$ y $EDAD$, ya que son los únicos campos de x mencionados en la consulta

- La respuesta a Q_1 es:

NOMBRE	EDAD
<i>luis</i>	20
<i>antonio</i>	23

Ejemplo: Consultas en CRT

■ Esquema:

- $ALUMNOS(ID, NOMBRE, EDAD, CIUDAD)$
- $CUR(IDC, NOMBREC)$
- $INS(ID, IDC, NOTA)$

- Q_2 : Encontrar el nombre de los alumnos, id del curso y nota obtenida por los alumnos

$$\{x \mid \exists y \in ALUMNOS \exists z \in INS (y.ID = z.ID \wedge x.NOMBRE = y.NOMBRE \wedge x.IDC = z.IDC \wedge x.NOTA = z.NOTA)\}$$

- La respuesta a Q_2 es:

NOMBRE	IDC	NOTA
<i>luis</i>	1	70
<i>luis</i>	2	85
<i>pedro</i>	2	80

Ejemplo: Consultas en CRT

- Esquema:

- $ALUMNOS(ID, NOMBRE, EDAD, CIUDAD)$
- $CUR(IDC, NOMBREC)$
- $INS(ID, IDC, NOTA)$

- Q_3 : Encontrar el nombre de los alumnos que inscribieron el curso con $IDC = 1$

$$\{x \mid \exists y \in ALUMNOS \exists z \in Ins(y.ID = z.ID \wedge z.IDC = 1 \wedge x.NOMBRE = y.NOMBRE)\}$$

- La respuesta a Q_3 es:

NOMBRE
<i>luis</i>

Ejemplo: Consultas en CRT

■ Esquema:

- $ALUMNOS(ID, NOMBRE, EDAD, CIUDAD)$
- $CUR(IDC, NOMBREC)$
- $INS(ID, IDC, NOTA)$

■ Q_4 : Encontrar el nombre de los alumnos que inscribieron el curso $BD2$

$$\{x \mid \exists y \in ALUMNOS \exists z \in INS \exists w \in CUR (y.ID = z.ID \wedge z.IDC = w.IDC \wedge w.NOMBREC = 'BD2' \wedge x.NOMBRE = y.NOMBRE)\}$$

■ La respuesta a Q_4 es:

NOMBRE
<i>luis</i>
<i>pedro</i>

Ejemplo: Consultas en CRT

- Esquema:

- $ALUMNOS(ID, NOMBRE, EDAD, CIUDAD)$
- $CUR(IDC, NOMBREC)$
- $INS(ID, IDC, NOTA)$

- Q_5 : Encontrar el nombre de los alumnos que inscribieron al menos dos cursos

$$\{x \mid \exists y \in ALUMNOS \exists z \in INS \exists w \in INS (y.ID = z.ID \wedge z.ID = w.ID \wedge z.IDC \neq w.IDC \wedge x.NOMBRE = y.NOMBRE)\}$$

- La respuesta a Q_5 es:

NOMBRE
<i>luis</i>

Ejemplo: Consultas en CRT

- Esquema:

- $ALUMNOS(ID, NOMBRE, EDAD, CIUDAD)$
- $CUR(IDC, NOMBREC)$
- $INS(ID, IDC, NOTA)$

- Q_6 : Encontrar el nombre de los alumnos que inscribieron todos los cursos

$$\{x \mid \exists y \in ALUMNOS \forall z \in CUR (\exists w \in INS (y.ID = w.ID \wedge w.IDC = z.IDC \wedge x.NOMBRE = y.NOMBRE))\}$$

- La respuesta a Q_6 es:

NOMBRE
<i>luis</i>

Cálculo Relacional de Dominios (CRD)

- Una variable de **dominio** es una variable que toma valores del dominio de algún atributo
- Una consulta en CRD es de la forma:

$$\{\langle x_1, \dots, x_n \rangle \mid p(x_1, \dots, x_n)\}$$

donde:

- cada x_i es una variable de dominio o una constante y
- $p(x_1, \dots, x_n)$ es una fórmula del CRD cuyas únicas variables libres son aquellas en x_1, \dots, x_n
- El resultado una consulta en CRD es el conjunto de todas las tuplas $\langle x_1, \dots, x_n \rangle$ para las que la fórmula $p(x_1, \dots, x_n)$ se evalúa como verdadera

Ejemplo: Consulta en CRD

- Dada la relación ALUMNOS:

ALUMNOS			
ID	NOMBRE	EDAD	CIUDAD
10	<i>luis</i>	20	<i>concepcion</i>
11	<i>pedro</i>	21	<i>chillan</i>
12	<i>antonio</i>	23	<i>concepcion</i>

- La consulta: “Listar el *ID* y *NOMBRE* de los alumnos” se expresa por:

$$\{\langle x, y \rangle \mid \exists z, w (\langle x, y, z, w \rangle \in Alumnos)\}$$

- El resultado de esta consulta es:

ID	NOMBRE
10	<i>luis</i>
11	<i>pedro</i>
12	<i>antonio</i>

Sintaxis de Consultas en CRD

- Una fórmula CRD se define de manera muy similar a una fórmula en CRT, pero ahora se trabaja con variables de **dominio**
- Dada una relación *Rel* con n atributos, las variables de tipo dominio X, Y y sea op uno de $\{<, >, =, \leq, \geq, \neq\}$
- Son fórmulas atómicas en CRD:
 - $\langle x_1, \dots, x_n \rangle \in Rel$ con cada x_i siendo una variable o una constante
 - $X op Y$
 - $X op c$ o $c op X$, con c constante

Fórmulas en CRD

- Una fórmula se define recursivamente de la siguiente manera:
 - Toda fórmula atómica p es una fórmula
 - Si p y q son fórmulas, también lo son: $\neg p$, $p \wedge q$, $p \vee q$, $p \rightarrow q$
 - $\exists X(p(X))$, donde X es una variable de dominio
 - $\forall X(p(X))$, donde X es una variable de dominio
- Los cuantificadores \exists, \forall limitan la variable X
- En una consulta CRD $\{\langle x_1, \dots, x_n \rangle \mid p(x_1, \dots, x_n)\}$, las variables x_i para $1 \leq i \leq n$ son las únicas variables libres

Ejemplo: Consultas en CRD

- Consideremos el siguiente esquema e instancia de BD:

CUR	
IDC	NOMBREC
1	<i>BD1</i>
2	<i>BD2</i>

INS		
ID	IDC	NOTA
10	1	70
10	2	85
11	2	80

ALUMNOS			
ID	NOMBRE	EDAD	CIUDAD
10	<i>luis</i>	20	<i>concepcion</i>
11	<i>pedro</i>	21	<i>chillan</i>
12	<i>antonio</i>	23	<i>concepcion</i>

Ejemplo: Consultas en CRD

- Esquema:

- $ALUMNOS(ID, NOMBRE, EDAD, CIUDAD)$
- $CUR(IDC, NOMBREC)$
- $INS(ID, IDC, NOTA)$

- Q_1 : Encontrar el nombre y la edad de los alumnos que viven en Concepción

$$\{\langle y, z \rangle \mid \exists xw (\langle x, y, z, w \rangle \in ALUMNOS \wedge w = 'concepcion')\}$$

- La consulta expresada en CRT es:

$$\{x \mid \exists y \in ALUMNOS (y.ciudad = 'concepcion' \wedge x.nombre = y.nombre \wedge x.edad = y.edad)\}$$

- La respuesta a Q_1 es:

NOMBRE	EDAD
<i>luis</i>	20
<i>antonio</i>	23

Ejemplo: Consultas en CRD

■ Esquema:

- $ALUMNOS(ID, NOMBRE, EDAD, CIUDAD)$
- $CUR(IDC, NOMBREC)$
- $INS(ID, IDC, NOTA)$

- Q_2 : Encontrar el nombre de los alumnos, id del curso y nota obtenida por los alumnos

$$\{\langle x_2, y_2, y_3 \rangle \mid \exists x_1 x_3 x_4 y_1 (\langle x_1, x_2, x_3, x_4 \rangle \in Alumnos \wedge \langle y_1, y_2, y_3 \rangle \in Ins \wedge x_1 = y_1)\}$$

- La respuesta a Q_2 es:

NOMBRE	IDC	NOTA
<i>luis</i>	1	70
<i>luis</i>	2	85
<i>pedro</i>	2	80

Ejemplo: Consultas en CRD

- Esquema:

- $ALUMNOS(ID, NOMBRE, EDAD, CIUDAD)$
- $CUR(IDC, NOMBREC)$
- $INS(ID, IDC, NOTA)$

- Q_3 : Encontrar el nombre de los alumnos que inscribieron el curso con $IDC = 1$

$$\{\langle x_2 \rangle \mid \exists x_1 x_3 x_4 y_1 y_2 y_3 (\langle x_1, x_2, x_3, x_4 \rangle \in Alumnos \wedge \langle y_1, y_2, y_3 \rangle \in Ins \wedge x_1 = y_1 \wedge y_2 = 1)\}$$

- La respuesta a Q_3 es:

NOMBRE
<i>luis</i>

Ejemplo: Consultas en CRD

- Esquema:

- $ALUMNOS(ID, NOMBRE, EDAD, CIUDAD)$
- $CUR(IDC, NOMBREC)$
- $INS(ID, IDC, NOTA)$

- Q_4 : Encontrar el nombre de los alumnos que inscribieron el curso $BD2$

$$\{\langle x_2 \rangle \mid \exists x_1 x_3 x_4 y_1 y_2 y_3 z_1 z_2 (\langle x_1, x_2, x_3, x_4 \rangle \in Alumnos \wedge \langle y_1, y_2, y_3 \rangle \in Ins \\ \wedge \langle z_1, z_2 \rangle \in Cur \wedge x_1 = y_1 \wedge y_2 = z_1 \wedge z_2 = 'BD2')\}$$

- La respuesta a Q_4 es:

NOMBRE
<i>luis</i>
<i>pedro</i>

Ejemplo: Consultas en CRD

- Esquema:

- $ALUMNOS(ID, NOMBRE, EDAD, CIUDAD)$
- $CUR(IDC, NOMBREC)$
- $INS(ID, IDC, NOTA)$

- Q_5 : Encontrar el nombre de los alumnos que inscribieron al menos dos cursos

$$\{\langle x_2 \rangle \mid \exists x_1 x_3 x_4 y_2 y_3 z_2 z_3 (\langle x_1, x_2, x_3, x_4 \rangle \in Alumnos \wedge \langle x_1, y_2, y_3 \rangle \in Ins \\ \wedge \langle x_1, z_2, z_3 \rangle \in Ins \wedge y_2 \neq z_2)\}$$

- La respuesta a Q_5 es:

NOMBRE
<i>luis</i>

Ejemplo: Consultas en CRD

- Esquema:

- $ALUMNOS(ID, NOMBRE, EDAD, CIUDAD)$
- $CUR(IDC, NOMBREC)$
- $INS(ID, IDC, NOTA)$

- Q_6 : Encontrar el nombre de los alumnos que inscribieron todos los cursos

$$\{ \langle x_2 \rangle \mid \exists x_1 x_3 x_4 (\langle x_1, x_2, x_3, x_4 \rangle \in Alumnos \wedge \forall z_1 z_2 \langle z_1, z_2 \rangle \in Cur \\ (\exists y_1 y_2 y_3 (\langle y_1, y_2, y_3 \rangle \in Ins \wedge y_1 = x_1 \wedge y_2 = z_1)))) \}$$

- La respuesta a Q_6 es:

NOMBRE
<i>luis</i>

Poder de Expresividad del AR y CR

- El AR y el CR tienen el mismo poder de expresividad
- Toda consulta expresada en AR puede ser expresada en CR
- ¿Qué pasa con las consultas expresadas en CR? ¿Pueden ser expresadas en AR?
- **Ejemplo:** ¿Qué expresa la siguiente consulta?

$$\{\langle x_1, x_2, x_3, x_4 \rangle \mid (\neg \langle x_1, x_2, x_3, x_4 \rangle \in \text{Alumnos})\}$$

Poder de Expresividad del AR y CR

- La consulta $\{\langle x_1, x_2, x_3, x_4 \rangle \mid (\neg \langle x_1, x_2, x_3, x_4 \rangle \in \text{Alumnos})\}$ es sintácticamente correcta
- Sin embargo, existe un problema, la consulta pregunta por todas las tuplas que no están en la tabla *ALUMNOS*
- El conjunto respuesta es entonces **infinito**
- Este tipo de consultas se denominan consultas **inseguras**

Consultas Seguras

- Consideremos una consulta Q y un conjunto I de instancias de relaciones, con una instancia por cada relación en Q
- Sea $Dom(Q, I)$ el conjunto de todas las constantes que aparecen en las relaciones en I o en la consulta Q
- Dado que I está restringido a ser finito, $Dom(Q, I)$ también es finito
- Una consulta en CR es **segura** si:
 - 1 Para cualquier conjunto I el conjunto respuesta para Q contiene solo valores que están en $Dom(Q, I)$
 - 2 Por cada $\exists X(p(X))$ en Q si encontramos valores para X que hacen verdadera la fórmula, entonces X contiene solo constantes en $Dom(Q, I)$
 - 3 Por cada $\forall X(p(X))$ en Q si asignamos valores para X que no están en $Dom(Q, I)$ entonces la fórmula debe ser verdadera

Consultas Seguras

- La consulta $\{\langle x_1, x_2, x_3, x_4 \rangle \mid (\neg \langle x_1, x_2, x_3, x_4 \rangle \in \text{Alumnos})\}$ no es segura porque:
 - La respuesta a esta consulta incluye valores que no están en $\text{Dom}(Q, I)$
- La consulta $\{\langle x_1, x_2 \rangle \mid \exists x_3, x_4 (\langle x_1, x_2, x_3, x_4 \rangle \in \text{Alumnos})\}$ es segura
- Toda consulta que puede ser expresada por medio de una **consulta segura en CR** puede también ser expresada en AR