Consultas conjuntivas

Clase 23

IIC 3413

Prof. Cristian Riveros

Outline

Consultas conjuntivas (CQ)

Evaluación de CQ

Optimización de CQ

Fragmento más sencillo: consultas conjuntivas

Definición

Una **consulta conjuntiva** (CQ) es una consulta en AR que solo contiene:

- proyección (π)
- selección sencilla ($\sigma_{A=B}$ o $\sigma_{A=v}$)
- Equality joins ($\bowtie_{A=B}$)
- Renaming ($\rho_{A \to B}$)

Ejemplo

```
SELECT   P.name, M.goals
FROM     Players AS P, Matches AS M, Players_Matches AS PM
WHERE    P.pId = PM.pId AND PM.mId = M.mId AND
         P.name = 'Alexi' AND M.year = 2001
```

En otras palabras, una consulta `SELECT-FROM-WHERE`.

Fragmento más sencillo: consultas conjuntivas

Definición

Una **consulta conjuntiva** (CQ) es una consulta en AR que solo contiene:

- proyección (π)
- selección sencilla ($\sigma_{A=B}$ o $\sigma_{A=v}$)
- Equality joins ($\bowtie_{A=B}$)
- Renaming ($\rho_{A \to B}$)

Sin perdida de generalidad

Desde ahora en adelante consideraremos consultas conjuntivas solo con:

- proyección π.
- selección $\sigma_{A=v}$.
- natural joins $\bowtie$.

$\sigma_{A=B}$, $\bowtie_{A=B}$ y $\rho_{A \to B}$ no cambian la complejidad del problema.

Fragmento más sencillo: consultas conjuntivas

Proposición

Para toda consulta conjuntiva Q, existe una consulta Q' tal que $Q(\mathcal{D}) = Q'(\mathcal{D})$ para toda BD $\mathcal{D}$ y Q' es de la forma:

$$\pi_l\big(\sigma_{c_1}(R_1) \bowtie \ldots \bowtie \sigma_{c_n}(R_n)\big)$$

con cada c_i una conjunción filtros $A = v$.

Demostración: use las reglas de reescritura.

Representación simplificada de consultas conjuntivas

Sea **V** un conjunto de variables y **C** un conjunto de constantes.

Simplificación

Desde ahora una consulta conjuntiva la representaremos como:

$$ans(\bar{y}) \ := \ R_1(\bar{x}_1),\, R_2(\bar{x}_2),\, \ldots,\, R_n(\bar{x}_n)$$

1. $\bar{x}_1, \ldots, \bar{x}_n$ son variables en **V** o constantes en **C**,
2. $\bar{y}$ es un subconjunto de variables en $\bar{x}_1, \ldots, \bar{x}_n$.

Ejemplo

$$ans(x, z) \ := \ P(x, \text{'Alexi'}),\, PM(x, y),\, M(y, 2001, z)$$

- x, y, z son variables.
- 'Alexi' y 2001 son constantes.

Representación simplificada de consultas conjuntivas

Sea **V** un conjunto de variables y **C** un conjunto de constantes.

Simplificación

Desde ahora una consulta conjuntiva la representaremos como:

$$ans(\bar{y}) \ := \ R_1(\bar{x}_1), R_2(\bar{x}_2), \ldots, R_n(\bar{x}_n)$$

1. $\bar{x}_1, \ldots, \bar{x}_n$ son variables en **V** o constantes en **C**,
2. $\bar{y}$ es un subconjunto de variables en $\bar{x}_1, \ldots, \bar{x}_n$.

Notación

- $R_1(\bar{x}_1), \ldots, R_n(\bar{x}_n)$ es el **cuerpo** de Q y $ans(\bar{y})$ es la **cabeza** de Q.
- cada $R_i(\bar{x}_i)$ es un **átomo** de Q.
- si $\bar{y}$ es **vacía**, entonces hablamos de una **consulta booleana**.

Homomorfismo de consultas conjuntivas

Definición

Un **homomorfismo** de Q a $\mathcal{D}$ es una función $h : (\mathbf{V} \cup \mathbf{C}) \to \mathbf{C}$ tal que:

- $h(c) = c$ para toda $c \in \mathbf{C}$ y
- si $R(d_1, \ldots, d_k)$ es un átomo de Q,

  entonces $(h(d_1), \ldots, h(d_k)) \in \mathcal{D}(R)$.

¿cuál es un homomorfismo de Q a $\mathcal{D}$?

$$Q : \quad anx(x, z) := P(x, \text{'Alexi'}, y), M(x, z, \text{'3'})$$

$\mathcal{D}$:

Players (P):

Id	Name	Year
1	Alexi	1987
2	Gary	1990
3	Arturo	1985

Matches (M):

Id	Stadium	Goals
1	Nacional	3
1	Monumental	3
2	San Carlos	4

Homomorfismo de consultas conjuntivas

Definición

Un **homomorfismo** de Q a $\mathcal{D}$ es una función $h : (\mathbf{V} \cup \mathbf{C}) \to \mathbf{C}$ tal que:

- $h(c) = c$ para toda $c \in \mathbf{C}$ y
- si $R(d_1, \ldots, d_k)$ es un átomo de Q,
  entonces $(h(d_1), \ldots, h(d_k)) \in \mathcal{D}(R)$.

Proposición

Para toda base de datos $\mathcal{D}$ y toda consulta conjuntiva Q de la forma:

$$ans(y_1, \ldots, y_k) \;:=\; R_1(\bar{x}_1),\, R_2(\bar{x}_2),\, \ldots,\, R_n(\bar{x}_n)$$

se tiene que $t \in Q(\mathcal{D})$ si, y solo si, existe un homomorfismo h de Q a $\mathcal{D}$ con

$$t = (h(y_1), \ldots, h(y_k)).$$

Demostración: ejercicio.

Outline

Consultas conjuntivas (CQ)

Evaluación de CQ

Optimización de CQ

¿qué tan complejo es evaluar una consulta conjuntiva?

Problema de **decisión**:

`PROBLEMA:`	Resultado no-vacío de consultas conjuntivas ($\textsc{CQ-Emptyness}$).
`INPUT:`	una consulta conjuntiva Q,
	una BD relacional $\mathcal{D}$
`OUTPUT:`	`TRUE` ssi $Q(\mathcal{D}) \neq \varnothing$.

Teorema

El problema $\textsc{CQ-Emptyness}$ es NP-completo.

Demostración: ejercicio.

¿cuáles son las consultas conjuntivas *difíciles*?

Ejemplo

Considere la siguiente consulta conjuntiva

$$R_1(A_1, A_2), R_2(A_2, A_3), \cdots, R_{n-1}(A_{n-1}, A_n), R_n(A_n, A_1)$$

con las siguientes relaciones:

R_i	A_i	A_{i+1}
	0	a
	0	b
	1	a
	1	b
	a	0
	a	1
	b	0
	b	1

R_n	A_n	A_1
	0	a
	0	b
	1	a
	1	b
	a	0
	a	1
	b	0
	b	1

- ¿cuál es el tamaño de sus relaciones intermedias?
- ¿cuál es el tamaño del resultado total?

¿estamos modelando el problema correctamente?

`PROBLEMA:` Resultado no-vacío de consultas conjuntivas (CQ-$\textsc{Emptyness}$).

`INPUT:` una consulta conjuntiva Q,

una BD relacional $\mathcal{D}$

`OUTPUT:` `TRUE` ssi $Q(\mathcal{D}) \neq \varnothing$.

En la práctica tenemos que:

$$|Q| \;<<\; |\mathcal{D}|$$

Consultas son muchísimo más pequeñas que los datos.

Complejidad en término de los datos

- **Combined**-complexity: consulta y datos son parte del input.
- **Data**-complexity: solo los datos son parte del input (consulta esta fija).

`PROBLEMA:`	Resultado no-vacío de consultas conjuntivas Q ($\mathrm{CQ}\text{-}\textsc{Eval}_Q$).
`INPUT:`	una BD relacional $\mathcal{D}$
`OUTPUT:`	$t \in Q(\mathcal{D})$.

Complejidad en término de los datos

Teorema

El problema $\textsc{ConjSQL-Eval}_Q$ esta en PTIME para todo consulta $Q \in \texttt{SQL}$.

¿es posible hacer una análisis mas fino?

Outline

Consultas conjuntivas (CQ)

Evaluación de CQ

Optimización de CQ

Equivalencia y satisfiabilidad de consultas conjuntivas

Definición

Un **homomorfismo** de Q_1 a Q_2 es una función $h : (\mathbf{V} \cup \mathbf{C}) \to (\mathbf{V} \cup \mathbf{C})$:

- $h(c) = c$ para toda $c \in \mathbf{C}$,
- si $R(d_1, \ldots, d_k)$ es un átomo de Q_1,
  entonces $R(h(d_1), \ldots, h(d_k))$ es un átomo de Q_2,
- si $ans(y_1, \ldots, y_k)$ es el cuerpo de Q_1,
  entonces $ans(h(y_1), \ldots, h(y_k))$ es el cuerpo de Q_2.

Proposición

Para todo par de consultas conjuntivas Q_1 y Q_2 se tiene que:

1. $Q_1(\mathcal{D}) \subseteq Q_2(\mathcal{D})$ para toda $\mathcal{D}$ si, y solo si,
2. existe un homomorfismo de Q_2 a Q_1.

Equivalencia y satisfiabilidad de consultas conjuntivas

`PROBLEMA:` Satisfabilidad de consultas conjuntivas. (CQ-SAT).

`INPUT:` una consulta conjuntiva Q,

`OUTPUT:` `TRUE` ssi existe $\mathcal{D}$ tal que $Q(\mathcal{D}) \neq \emptyset$.

`PROBLEMA:` Igualdad de consultas conjuntivas $(\text{CQ-}\textsc{Equivalence})$.

`INPUT:` consultas conjuntivas Q_1 y Q_2,

`OUTPUT:` `TRUE` ssi para todo $\mathcal{D}$ se cumple $Q_1(\mathcal{D}) = Q_2(\mathcal{D})$.

Teorema

- CQ-SAT es un problema trivial (siempre es satisfacible).
- $\text{CQ-}\textsc{Equivalence}$ es $\text{NP-}\textsc{completo}$.