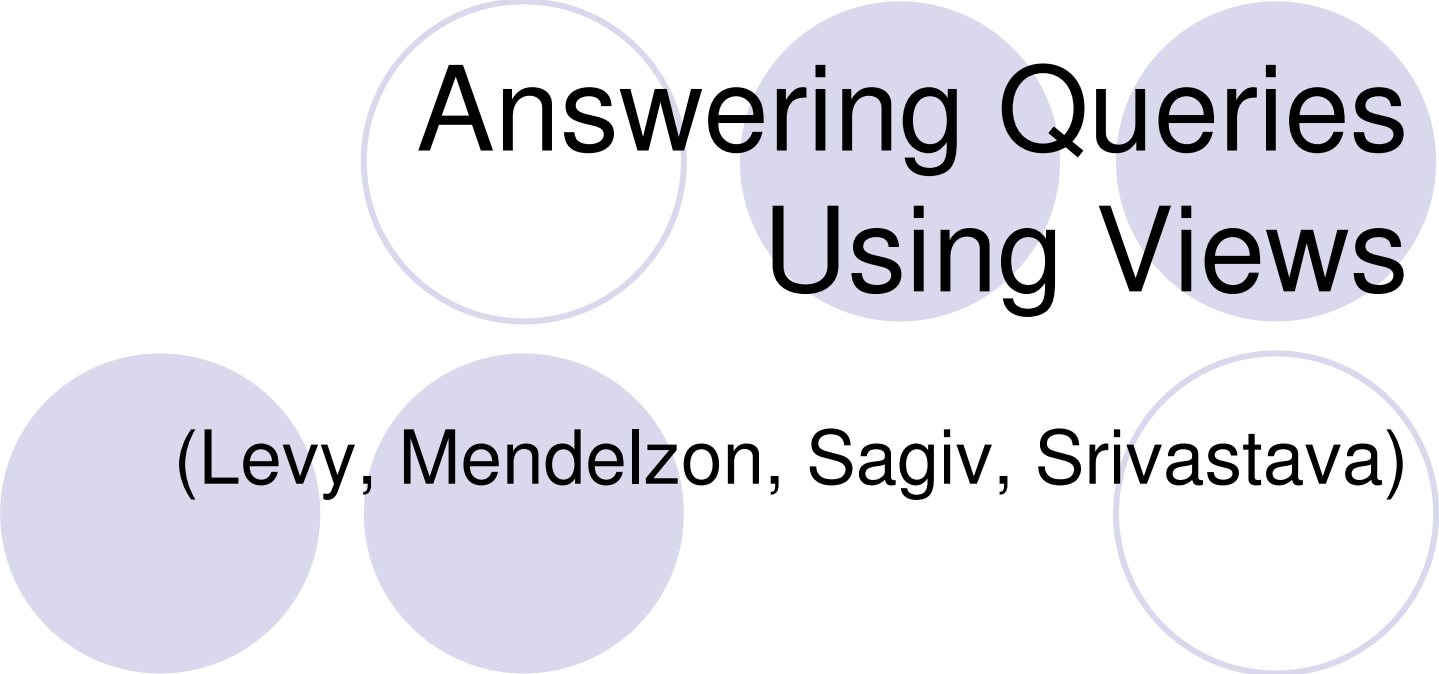




Answering Queries Using Views

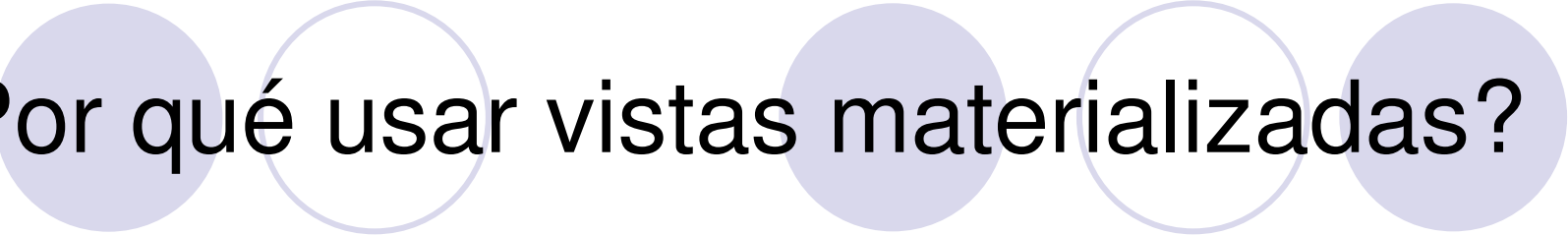
(Levy, Mendelzon, Sagiv, Srivastava)

Javier López C.

Abstract



- Se analizará el problema de responder consultas usando vistas materializadas
- Nos enfocamos en la reescritura de consultas usando las vistas
- Análisis de complejidad



¿Por qué usar vistas materializadas?

- Mejorar la eficiencia en tiempos de respuesta
- Relaciones “virtuales” en consultas
- Fuentes no disponibles
- Alto costo en acceder a fuentes

Reescrituras

- Una consulta Q' es una *reescritura* de Q que usa las vistas $V=V_1, \dots, V_m$ si:
 - Q y Q' son equivalentes (producen el mismo resultado para cualquier base de datos), y
 - Q' contiene uno o más literales de V
- Q' es *localmente mínima* si no se pueden eliminar más literales sin perder la equivalencia con Q
- Q' es *globalmente mínima* si no existe otra reescritura con menos literales

Ejemplo

- Q: $q(X,U) \text{ :- } p(X,Y), p_0(Y,Z), p_1(X,W), p_2(W,U)$
- V: $v(A,B) \text{ :- } p(A,C), p_0(C,B), p_1(A,D)$
- Q': $q(X,U) \text{ :- } v(X,Z), p_1(X,W), p_2(W,U)$

Otro ejemplo

Al ejemplo anterior le agregamos

- $v_1(A) \text{ :- } p(A,C), p_1(A,D)$
- $q(X,U) \text{ :- } v_1(X), p(X,Y), p_0(Y,Z), p_1(X,W),$
 $p_2(W,U)$

En este paper nos enfocamos a reducir los literales, aunque no necesariamente siempre es más óptimo

Reescrituras completas

- Q' es una *reescritura completa* de Q usando $V = V_1, \dots, V_m$ si es una reescritura que sólo contiene literales de V y predicados integrados

Ejemplo

- $v_2(A,B) \text{ :- } p_1(A,C), p_2(C,B), p_0(D,E)$
- $Q'': q(X,U) \text{ :- } v(X,Z), v_2(X,U)$

Mapeos de contención

- Un mapeo de contención de Q_1 a Q_2 es un mapeo de las variables de Q_1 a las de Q_2 tal que cada literal de Q_1 esté mapeado a uno de Q_2
- En el primer ejemplo existe un mapeo $\{A \rightarrow X, B \rightarrow Z, C \rightarrow Y, D \rightarrow W\}$

Mapeos de contención

- Si ni Q_1 ni Q_2 contienen literales integrados, o sólo Q_2 los tiene, encontrar un mapeo es condición necesaria y suficiente para demostrar contención
- Si Q_1 tiene literales integrados, es sólo una condición suficiente

Mapeos de contención

- En el caso anterior, si Q_1 contiene a Q_2 , existen $Q_2^1, \dots, Q_2^n$ tal que:
 - $Q_2^1, \dots, Q_2^n$ difieren sólo en sus literales integrados,
 - Q_2 es equivalente a la unión de $Q_2^1, \dots, Q_2^n$,
 - Para todo i , $1 \leq i \leq n$, hay un mapeo φ_i de Q_1 a Q_2^i tal que $bi(Q_2^i)$ conlleve $\varphi_i(bi(Q_1))$, donde $bi(Q)$ es la conjunción de átomos integrados de la consulta Q

Ejemplo

○ $Q_1: q(Y) :- e(Y), r(U_1, V_1), U_1 \leq V_1$

○ $Q_2: q(X) :- e(X), r(U, V), r(V, U)$

● Q_2 es la unión de:

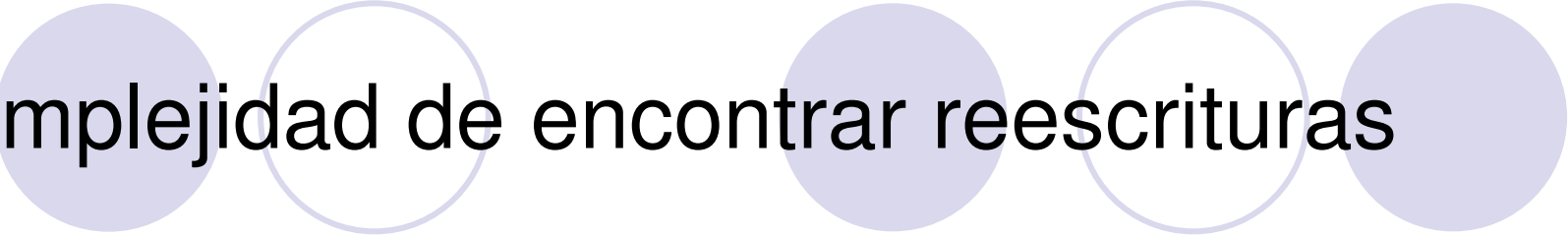
○ $Q_2^1: q(X) :- e(X), r(U, V), r(V, U), U \leq V$

○ $Q_2^2: q(X) :- e(X), r(U, V), r(V, U), V \leq U$

● Y los mapeos serían:

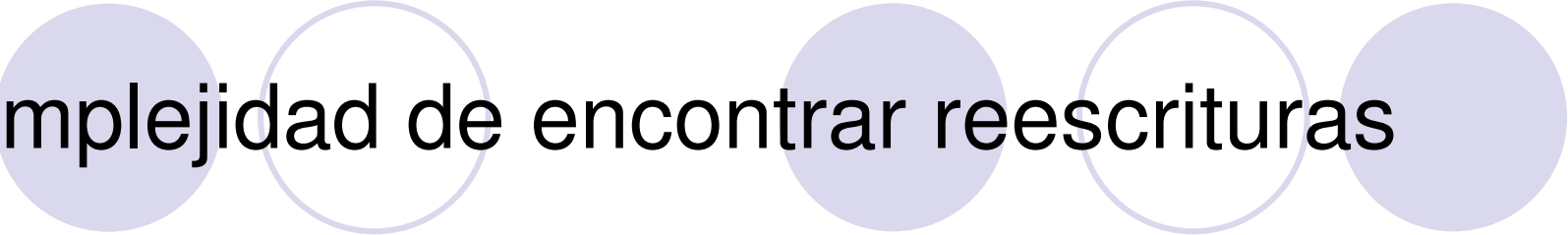
○ $\varphi_1: \{Y \rightarrow X, U_1 \rightarrow U, V_1 \rightarrow V\}$

○ $\varphi_2: \{Y \rightarrow X, U_1 \rightarrow V, V_1 \rightarrow U\}$



Complejidad de encontrar reescrituras

- Sean Q y V consultas conjuntivas con predicados integrados. Existe una reescritura de Q usando V ssi $\pi_{\emptyset}(Q) \subseteq \pi_{\emptyset}(V)$, o sea, si V es vacío para una base de datos, también lo es Q



Complejidad de encontrar reescrituras

- Sean Q una consulta y V un conjunto de vistas:
 - Si Q es una consulta conjuntiva con predicados integrados, y V son vistas conjuntivas sin predicados integrados, determinar si existe una reescritura de Q usando V es NP-completo
 - Si Q y V son conjuntivas con predicados integrados, este problema es Π^P_2 -completo

Complejidad de encontrar reescrituras

- Sea Q una consulta conjuntiva sin predicados integrados
 $q(\bar{X}) \text{ :- } p_1(\bar{U}_1), \dots, p_n(\bar{U}_n)$
y V un conjunto de vistas sin predicados integrados
 - 1) Si Q' es una reescritura localmente mínima de Q usando V , el conjunto de literales de relaciones de la base de datos en Q' es isomorfo a un subconjunto de los literales en Q

Complejidad de encontrar reescrituras

○ 2) Si

$$q(\bar{X}) \text{ :- } p_1(\bar{U}_1), \dots, p_n(\bar{U}_n), v_1(\bar{Y}_1), \dots, v_k(\bar{Y}_k)$$

es una reescritura de Q usando V , existe

$$q(\bar{X}) \text{ :- } p_1(\bar{U}_1), \dots, p_n(\bar{U}_n), v_1(\bar{Y}'_1), \dots, v_k(\bar{Y}'_k)$$

donde $\{\bar{Y}'_1, \dots, \bar{Y}'_k\} \subseteq \{\bar{U}_1, \dots, \bar{U}_n\}$, es decir, hay una reescritura que no introduce nuevas variables

○ 3) Si Q y V incluyen predicados integrados, existe una reescritura como la anterior, pero con la diferencia de que la reescritura puede ser una unión de consultas conjuntivas

Ejemplo

- Q: $q(X,Y,U,W) \text{ :- } p(X,Y), r(U,W), r(W,U)$
V: $v(A,B,C,D) \text{ :- } p(A,B), r(C,D), C \leq D$
V tiene predicado integrado, 2) no se aplica
- Q': $q(X,Y,U,W) \text{ :- } v(X,Y,C,D), r(U,W), r(W,U)$
No se puede hacer una reescritura conjuntiva sin agregar variables, pero sí disyuntiva:
- Q': $q(X,Y,U,W) \text{ :- } v(X,Y,C,D), r(U,W)$
Q': $q(X,Y,U,W) \text{ :- } v(X,Y,C,D), r(W,U)$

Reescrituras mínimas

- Sea Q una consulta conjuntiva y V un conjunto de vistas, ambos sin predicados integrados. Si el cuerpo de Q tiene p literales y Q' es una reescritura localmente mínima y completa de Q usando V , Q' tiene como máximo p literales

Ejemplo

- Esta propiedad no se mantiene si existen dependencias funcionales
- $q(X,Y,Z) \text{ :- } e(X,Y,Z)$
 $v_1(X,Y) \text{ :- } e(X,Y,Z)$
 $v_2(X,Z) \text{ :- } e(X,Y,Z)$
- Si en e , X determina a Y y Z , la única reescritura completa sería:
- $q(X,Y,Z) \text{ :- } v_1(X,Y), v_2(X,Z)$

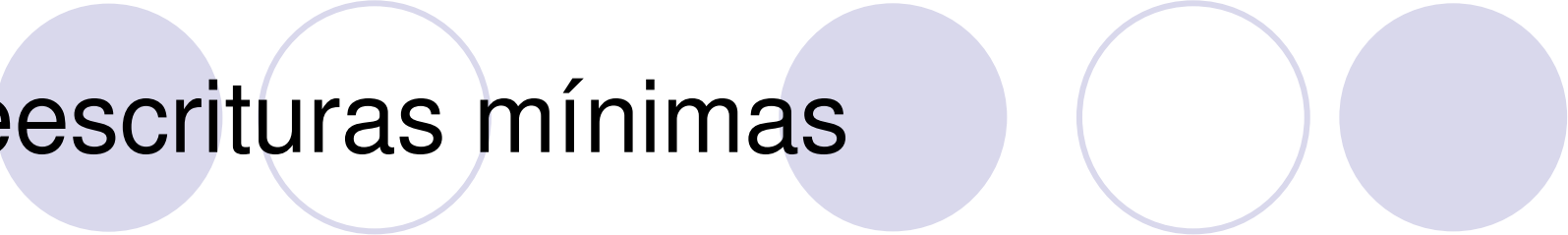
Reescrituras mínimas

- Sean Q y V conjuntivas sin predicados integrados
 - 1) El problema de si existe una reescritura Q' de Q usando V tal que Q' tiene no más de k literales, con k menor o igual al número de literales del cuerpo de Q , es NP-completo
 - 2) Lo anterior pero con no más de k literales de relaciones de bases de datos, también es NP-completo
 - 3) Si existe una reescritura completa de Q usando V , también es NP-completo
 - 4) Si Q y V tienen predicados integrados, 1) es Σ^P_3

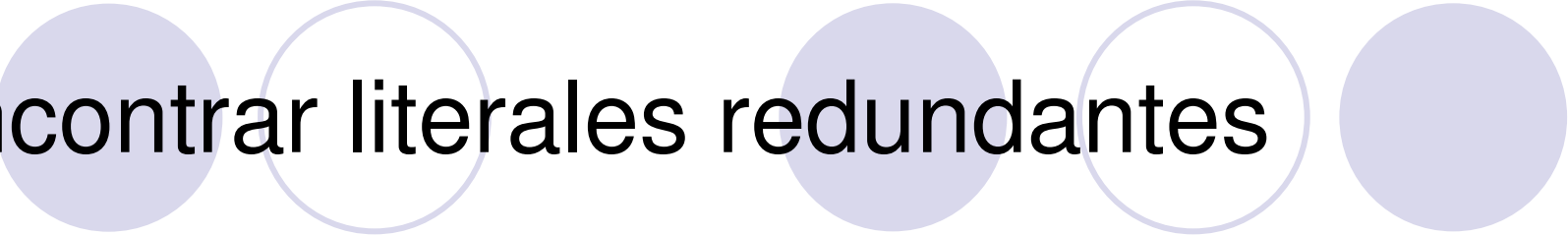
Reescrituras mínimas

- El problema de encontrar una reescritura mínima global para Q usando V , con Q y V conjuntivas sin predicados integrados, es Σ^p_2
- El problema de si existe una reescritura completa para Q usando V , con Q y V disyunciones de conjunciones sin predicados integrados, es NP-completo

Reescrituras mínimas



- El problema de la reescritura completa es NP-completo para consultas conjuntivas y vistas sin predicados integrados incluso cuando ambas estén definidas por reglas que no contengan predicados repetidos en sus cuerpos



Encontrar literales redundantes

- Para encontrar una reescritura se deben hacer 2 pasos:
 - aplicar mapeos de las vistas a la consulta y agregar los literales correspondientes, y
 - eliminar los literales originales que sean redundantes en la nueva consulta
- En general, cada uno de estos pasos contribuye con complejidad exponencial

Algoritmo

- 1) $Q: q(\bar{X}) :- p_1(\bar{U}_1), \dots, p_n(\bar{U}_n)$
- 2) $V: v(\bar{Z}) :- r_1(\bar{W}_1), \dots, r_m(\bar{W}_m)$
- Sea h un mapeo de contención del cuerpo de v al de q , y sea el resultado:
- 3) $q(\bar{X}) :- p_1(\bar{U}_1), \dots, p_n(\bar{U}_n), v(\bar{Y})$
donde $\bar{Y} = h(\bar{Z})$
- Reemplazamos $v(\bar{Y})$ por el cuerpo de 2), reemplazando cada variable T que aparece en $\bar{Z}$ por $h(T)$ y las que no aparecen por variables nuevas
- 4) $q(\bar{X}) :- p_1(\bar{U}_1), \dots, p_n(\bar{U}_n), r_1(\bar{V}_1), \dots, r_m(\bar{V}_m)$
- 5) Se eliminan los literales no necesarios según las siguientes definiciones

Definiciones

- Un literal $r_j(V_j)$ cubre a $p_i(U_i)$ que tiene el mismo predicado si se cumple que:
 - Si $p_i(U_i)$ tiene una variable en X o constante, $r_j(V_j)$ también la tiene en la misma posición
 - Si los argumentos en las posiciones a_1 y a_2 de $p_i(U_i)$ son iguales, lo mismo ocurre en $r_j(V_j)$

Definiciones



- El conjunto de literales necesarios N es el mínimo subconjunto de literales de Q que satisface alguna de las siguientes 4 condiciones:
 - 1) Literales que no tienen asociado
 - 2) Literales que no son cubiertos por sus asociados

Definiciones

- 3) Si las siguientes 5 condiciones se satisfacen:
 - El literal $p_i(U_i)$ tiene la variable T en el argumento a_1
 - El asociado de $p_i(U_i)$ tiene la variable H en el argumento a_1
 - La variable H no está en Y
 - T también aparece en el argumento a_2 de $p_l(U_l)$
 - El asociado de $p_l(U_l)$ no tiene H en a_2
- 4) Si $p_i(U_i)$ está en N y la variable T aparece en $p_i(U_i)$, si $p_l(U_l)$ tiene T en el argumento a y su asociado no, $p_l(U_l)$ también está en N

Ejemplo

- Siguiendo con el primer ejemplo, al aplicar los primeros 4 pasos se obtiene:
- $q(X,U) \text{ :- } p(X,Y), p_0(Y,Z), p_1(X,W), p_2(W,U), p(X,C), p_0(C,Z), p_1(X,D)$
- De los originales, p_2 es necesario por 1), y p_1 es necesario por 4), con lo que queda:
- $q(X,U) \text{ :- } p_1(X,W), p_2(W,U), v(X,Z)$

Teorema



- 1) La consulta $q(X) :- N, v(Y)$ es una reescritura de Q usando V
- 2) Supongamos que h no mapea 2 literales $r_j(V_j)$ al mismo literal en 1), y 1) es mínima. Entonces el conjunto maximal de $p_i(U_i)$ redundantes es único y es exactamente el complemento de N

Predicados integrados



- Se puede aplicar la misma base del algoritmo anterior, pero como vimos al principio, el resultado cuando hay predicados integrados es una disyunción de consultas $Q_1, \dots, Q_m$ que difieren sólo en los predicados integrados
- Teniendo los m mapeos, sólo podemos eliminar los literales que no sean necesarios para ningún mapeo

Ejemplo

- $Q: q(X,Y,U,W) :- p(X,Y), r(U,W), r(W,U)$
- $V: v(A,B,C,D) :- p(A,B), r(C,D), C \leq D$
- $Q': q(X,Y,U,W) :- p(X,Y), r(U,W), r(W,U), v(A,B,C,D)$
- Q puede ser reescrito como la unión de:
- $Q_1: q(X,Y,U,W) :- p(X,Y), r(U,W), r(W,U), U \leq W$
- $Q_2: q(X,Y,U,W) :- p(X,Y), r(U,W), r(W,U), W \leq U$
- Los mapeos de Q' a Q_1 y Q_2 son:
- $\phi_1: \{X \rightarrow X, Y \rightarrow Y, U \rightarrow U, W \rightarrow W, C \rightarrow U, D \rightarrow W\}$
- $\phi_2: \{X \rightarrow X, Y \rightarrow Y, U \rightarrow U, W \rightarrow W, C \rightarrow W, D \rightarrow U\}$
- Para ϕ_1 sólo $r(W,U)$ es necesario, y para ϕ_2 sólo $r(U,W)$, por lo que $p(X,Y)$ es el único que no es necesario en ninguno de los 2
- $q(X,Y,U,W) :- r(U,W), r(W,U), v(X,Y,C,D)$