

Resumen de teórico Base de Datos

Agustin Curto, agucurto95@gmail.com

2016

Índice general

1. Modelado Entidad - Relación	2
1.1. Conjunto de Entidades	2
1.2. Claves	2
1.3. Correspondencia de cardinales	3
1.4. Formas de participación	3
1.5. Conjunto de entidades débiles	3
1.6. Diseño de una DB de calidad	4
1.6.1. Eliminación de información redundante	4
1.6.2. Uso de CE vs Atributos	4
1.6.3. Errores comunes	4
1.6.4. Uso de CEs vs uso de CR	5
1.6.5. CR de grado ≥ 2 vs CR binarias	5
1.6.6. CE débiles vs Atributo multivalorado compuesto	5
1.7. Especialización - Generalización	5
1.8. Esquemas e instancias relacionales	6
1.9. Tipos de atributos	7
1.10. DB relacionales	7
1.10.1. Superclaves	7
1.10.2. Claves candidatas y primarias	7
1.10.3. Claves foráneas	8
1.11. Diseño de una DB relacional	8
1.12. Reducción a esquemas relacionales	8
1.12.1. Regla CEF1	9
1.12.2. Regla CEF2	9
1.12.3. Regla CEF3	9
1.12.4. Regla CR1	10
1.12.5. Regla CR2	10
1.12.6. Regla CR3	11
1.12.7. Regla CED	11
1.12.8. Reducción de generalización	11
2. Algebra Relacional	14

Capítulo 1

Modelado Entidad - Relación

1.1. Conjunto de Entidades

- Una **entidad** es un objeto que existe y es distinguible de los otros objetos. Las entidades tienen atributos.

Ejemplo: una persona tiene nombres y direcciones.

- Un **conjunto de entidades** es un conjunto de entidades del mismo tipo (i.e. Con los mismos atributos) que comparte las mismas propiedades.

Ejemplo: conjunto de todas las personas con los atributos del ejemplo anterior.

- Un **dominio** es el conjunto de valores permitidos para cada atributo.
- Tipos de atributos:
 - Atributos Simples y compuestos.
 - Atributos uni-valorados y multi-valorados
 - Atributos derivados, pueden computarse de otros atributos.

Notaciones de diagrama

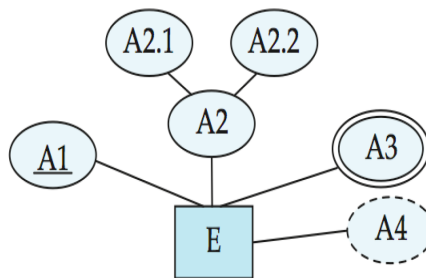


Figura 1.1: Entidad de nombre E, Atributo simple A1, Atributo compuesto A2, Atributo multivalorado A3, Atributo derivado A4, Clave primaria A1

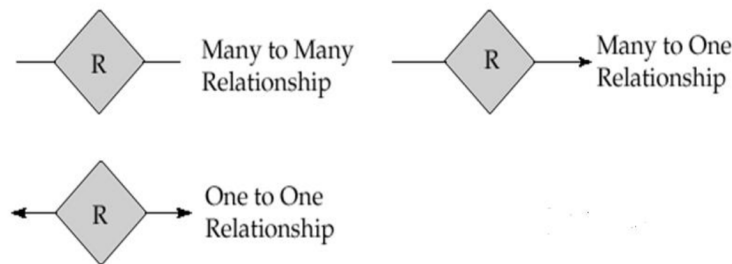
1.2. Claves

- Una **superclave** de un CE es un conjunto de uno o más atributos cuyos valores únicamente determinan cada entidad.

- Una **clave candidata** de un CE es una superclave minimal (i.e. si se quita atributo dejamos de tener superclave).
- Una **clave primaria** es aquella seleccionada de entre todas las claves candidatas que puedan existir.

1.3. Correspondencia de cardinales

- **Conjuntos de relaciones uno-uno:** una entidad de **E1** está asociada con a lo más una entidad de **E2** vía una relación **R**. Una entidad de **E2** está asociada con a lo más una entidad de **E1** vía **R**.
- **Conjuntos de relaciones uno-varios / varios-uno:** una entidad de **E1** está asociada con ninguna o varias entidades de **E2** vía una relación **R**. Una entidad de **E2** está asociada con a lo más una entidad de **E1** vía **R**.
- **Conjuntos de relaciones varios-varios:** una entidad de **E1** está asociada con ninguna o varias entidades de **E2** vía **R**. Una entidad de **E2** está asociada con a lo más una entidad de **E1** vía **R**.



1.4. Formas de participación

- **Participación total:** toda entidad en el conjunto de entidades participa en al menos una relación en el conjunto de relaciones. Es indicada por una línea doble.

Ejemplo: participación de caja de ahorro en cliente es total. Toda caja de ahorro debe tener clientes asociados.

- **Participación parcial:** algunas entidades no participan en alguna relación en el conjunto de relaciones.

Ejemplo: participación de instructor en supervisor es parcial.

1.5. Conjunto de entidades débiles

Un CE que no tiene una clave primaria en el conjunto de sus atributos, se llama **conjunto de entidades débiles**. La existencia de un CE débiles depende de la existencia de un CE fuertes llamado **CE identificador**.

Ejemplo: ¿En el caso de libro-biblioteca cuál sería el CE identificador?

Hay un CR varios-uno entre CE débil y CE identificador, donde el CE débil tiene participación total, a este CR se le llama **CR de identificación**. El mismo se representa con un diamante doble.

El **discriminador** de un CE débiles es un conjunto de atributos que permite distinguir entre todas las entidades de un CE débiles asociadas a la misma entidad fuerte.

Para identificar las entidades débiles se forma la clave primaria del CE con la clave primaria del CE identificador más el discriminador del CE débiles.

1.6. Diseño de una DB de calidad

1.6.1. Eliminacion de información redundante

Una vez que se han decidido los CE del problema y los atributos de los CE se pasa a definir los CR, esos CR pueden señalar que algunos atributos en CEs son redundantes y por lo tanto necesitan ser removidos.

Ejemplo:

- CE **aula** con atributos (edificio, número y capacidad)
- CE **sección** con atributos (idCurso, idSec, semestre, año, edificio, número)
- CR **sec-aula** que relaciona secciones con aulas

¿Qué pasa con esta propuesta?

- edificio y número aparecen en ambos CE pero son clave primaria de aula
- Luego edificio y número son redundantes en **sección** y deben ser removidos de allí

1.6.2. Uso de CE vs Atributos

A veces tenemos dos alternativas para modelar un objeto perteneciente a un CE E, como un CE o como un atributo de E.

- Cuando el CE E tiene un solo objeto que tiene un solo valor (atómico), conviene modelar al objeto como un atributo simple.
- Cuando el CE E tiene cero o más objetos que tienen un solo valor (atómico), se puede modelar al objeto como un atributo multivalorado.
- Cuando el objeto tiene varios atributos y es poseído por una sola entidad del CE E, conviene modelar el objeto como CE. La razón es que “ser poseído por una sola entidad” no se puede representar usando un atributo.
- Cuando el objeto tiene varios atributos algunos de los cuales no son de clave primaria del objeto y es poseído por varias entidades del CE E, conviene modelar el objeto como CE. La razón es que los atributos del objeto que no se usan para identificarlo generan redundancia de información, cosa que no pasa cuando el objeto se representa un como CE.

1.6.3. Errores comunes

Un error muy común es usar la clave primaria de un CE como atributo de otro CE en lugar de usar un CR. La forma correcta sería usar un CR que vincule los dos CE, ya que hace la conexión entre los dos CE explícita, en lugar de implícita vía atributos de la clave primaria de uno de los CE.

1.6.4. Uso de CEs vs uso de CR

Muchas veces son posibles dos alternativas para modelar un asunto, como un CE o como un CR.

- Si CR es uno-varios o varios-uno: si hay atributos, no de clave primaria en el CE A, entonces los atributos de A que no son de clave primaria dan lugar a redundancia de información si se usa alternativa CE. Por lo que conviene usar la opción CR en lugar de CE.
- Si el CR es varios – varios: si en al menos un CE hay atributos que no son de clave primaria, entonces ocurre igual que antes, los atributos del CE que no son de clave primaria dan lugar a redundancia de información si se usa la opción CE. Por lo que conviene usar CR en lugar de CE.

1.6.5. CR de grado ¿2 vs CR binarias

A veces tenemos las alternativas de modelar una situación como un CR de grado ¿2 o como varios CR binarios. Si con los CR binarios puedo capturar los datos del CR de grado ¿2 y puedo expresar más restricciones de integridad que con el CR de grado ¿2, entonces usar los CR binarios.

1.6.6. CE débiles vs Atributo multivalorado compuesto

A veces para modelar una situación tenemos dos alternativas un CE débiles o un atributo multi-valorado compuesto. Si el objeto contemplado tiene su complejidad e involucra varios atributos y es de suma importancia para la organización, es poco conveniente tenerla como un atributo de un CE fuerte, por lo tanto, conviene modelarlo como un CE débiles.

Si los objetos siendo considerados están relacionados con entidades de un CE fuerte que no es de identificación, entonces es necesario modelar dichos objetos como CE débiles.

1.7. Especialización - Generalización

Cuando en un diseño ER hay varios CE que son bastante similares en el sentido que comparten varios atributos en común, que tienen las mismas claves primarias y que participan en los mismos CR, aparecen problemas de repetición.

Esto es malo ya que estos modelos ocupan demasiado espacio. Muchos CR hacen el diagrama más intrincado, esto se ve agravado cuando el esquema de la BD tiene muchos CE y CR. Al cambiar un CE o CR muchas veces hay que propagar el cambio a otros CE o CRs.

La solución a este problema es utilizar especialización-generalización.

- **Especialización:** hace referencia a un proceso de diseño *top-down* donde designamos subgrupos dentro de un CE que son distintivos de otras entidades en el CE. Estos subgrupos son CE de más bajo nivel que tienen atributos específicos o participan de CR que no aplican al CE de más alto nivel. Una especialización se denota con un triángulo etiquetado ISA. La relación ISA se llama también relación de superclase – subclase.
- **Herencia de atributos:** un CE de más bajo nivel hereda:
 - todos los atributos
 - la clave primaria

- participaciones en CR del CE de más alto nivel con el cual está relacionado.
- **Generalización:** hace referencia a un proceso de diseño *bottom up* que generaliza unos cuantos CE que comparten las mismas propiedades en un CE de más alto nivel.

Restricciones de integridad:

Para indicar si una entidad pertenece o no a más de un CE de nivel más bajo dentro de la generalización.

- Disjunto: una entidad puede pertenecer a solo un CE de nivel más bajo. Usar palabra reservada *disj*.
- Solapado: una entidad puede pertenecer a más de un CE de nivel más bajo.

Restricciones de completitud:

Para indicar si una entidad en el CE de nivel más alto debe pertenecer a al menos uno de los CE de nivel más bajo en la generalización.

- Total: una entidad debe pertenecer a un CE de nivel más bajo. Usar línea doble.
- Parcial: una entidad puede no pertenecer a un CE de nivel más bajo.

1.8. Esquemas e instancias relacionales

Definiciones:

- A_1, A_2, \dots, A_n son **atributos**
- $R = (A_1, A_2, \dots, A_n)$ es un **esquema de relación**
Dado un enunciado en lenguaje natural, se puede identificar atributos primero, y luego se pueden proponer esquemas de relación para distintos conceptos del problema actual.
- Dados conjuntos D_1, D_2, \dots, D_n una **relación r** es un subconjunto de D_1, D_2, \dots, D_n . Las relaciones son conjuntos (el orden entre las tuplas no importa).
- Se puede pensar una relación como una tabla con columnas para los dominios D_1, D_2, \dots, D_n respectivamente.
- Un **elemento** $t \in r$ es una tupla representado mediante una fila de la tabla.

Notar que el concepto de relación en el modelo relacional no tiene nada que ver con el concepto de relación en el modelo ER. Es más bien como están acostumbrados a trabajar en matemática.

Se puede expresar la información manejada por una organización como un conjunto de relaciones o tablas, en cada columna se deciden los dominios con los que se trabaja para los valores.

Notación: $r(R)$ significa r es una relación con esquema de relación R , es decir, las columnas de r tienen como nombres los atributos de R .

- Los conceptos del problema que se asociaban a esquemas de relación ahora se terminan asociando también a relaciones o tablas de la BD.

- Los dominios de las tablas pasan a ser los dominios de los atributos del esquema relacional.
- Además es más fácil de comprender una tabla donde las columnas tienen los nombres de un esquema de relación.

1.9. Tipos de atributos

El conjunto de valores permitidos para cada atributo se llama **dominio del atributo**. Se requiere que los valores de los atributos sean atómicos; esto es, indivisibles.

A veces no sabemos el valor de un atributo o el mismo aún no existe, para poder reflejar esto en la base de datos se utiliza el valor especial **null**, el cual es un miembro de todo dominio y significa que el valor es desconocido o no existe. Si para una tupla no tenemos el valor de un atributo por algún motivo, podemos poner null como valor para ese atributo.

1.10. DB relacionales

Definiciones:

- Un **esquema de una BD relacional** es un conjunto de esquemas de relación $R_1, R_2 \dots R_n$
- Una **base de datos relacional** consiste de múltiples relaciones $r_1(R_1), r_2(R_2) \dots r_n(R_n)$
Primero definimos los conceptos del problema mediante esquemas de relación, luego definimos los datos como tablas asociadas a esos esquemas, por último podemos poblar las tablas, consultarlas y alterarlas.
- nombre de relación, expresado en minúscula, r, s, u, r_1, r_2, \dots
- nombre de esquema de relación, expresado en mayúscula R, S, U, R_1, R_2, \dots
- sea $t \in r, r(R), A \in R$. $t[A]$ es el valor de t en A .
- sea $t \in r, r(R)$. $t[i]$ es el valor de t en atributo i -ésimo de R .

1.10.1. Superclaves

Definición: Sea $K \in R$, R esquema de relación; K es una **superclave** de R si los valores para K son suficientes para identificar una tupla única en cada posible relación $r(R)$. Observar que se habla de relaciones para el problema o situación del mundo real que está siendo considerado.

1.10.2. Claves candidatas y primarias

Una superclave K es una **clave candidata** si K es minimal, es decir, para todo atributo de K si se lo quito dejo de tener una superclave.

No confundir clave candidata con superclave de cardinalidad mínima. Una de las claves candidatas es elegida para ser la clave primaria.

Notación: Se indican los atributos de una clave primaria para un esquema de relación R subrayando los atributos de R que forman la clave primaria.

1.10.3. Claves foráneas

Sea R_1 una **relación referenciante** y R_2 una **relación referenciada**, entonces

- Restricción de clave foránea: Los valores de uno o más atributos en una tupla de R_1 aparecen en uno o más atributos de una tupla en R_2 .
- Los atributos referenciados en R_1 suelen formar una clave primaria del esquema de R_2 .
- Los atributos referenciados de R_1 suelen formar una clave candidata del esquema de R_2 .

Representación gráfica:

- Los esquemas se representan con rectángulos conteniendo nombre de esquema y nombres de atributos.
- Los atributos de clave primaria se subrayan.
- Las restricciones de clave foránea se representan mediante flechas que van de atributos referenciantes (de esquema referenciante) a atributos referenciados (de esquema referenciado).

1.11. Diseño de una DB relacional

Existen dos formas de diseñar una buena base de datos:

- Aprovechar un buen diseño de esquema ER: si se tiene un buen diseño de ER, se lo puede transicionar a un buen diseño relacional.
- Obtener un buen diseño relacional a partir de los atributos y las restricciones de integridad del problema. La **teoría de normalización** trabaja con esta idea y trata con cómo diseñar buenos esquemas de BD relacionales.

1.12. Reducción a esquemas relacionales

Se tiene un buen diseño de esquema de BD usando modelado-ER y se va a traducir a un conjunto esquemas relacionales con claves primarias y restricciones de clave foránea. Hay reglas de traducción que suelen resolver un porcentaje alto del problema pero esas reglas no contemplan todos los casos.

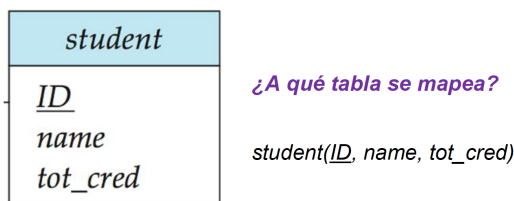
Una propiedad de la traducción de diagramas ER a tablas que se debe respetar es que para cada CE y CR hay un esquema relacional único al cual se asigna el CE / CR.

Además de los esquemas de relación obtenidos por la traducción automática se debe obtener:

- Claves primarias de esquemas de relación: para esto se usa la información de claves primarias de CE.
- Restricciones de clave foránea: para esto se usa el tipo de elemento de modelo ER a mapear (p.ej. CR, CE débiles, atributo multivalorado, etc.).

1.12.1. Regla CEF1

Un CE fuerte que no involucra atributos compuestos ni atributos multivalorados se mapea a un esquema relacional con los mismos atributos. La clave primaria del CE se convierte en la clave primaria del esquema relacional.



1.12.2. Regla CEF2

Un CE fuerte que no involucra atributos/subatributos multivalorados se mapea a un esquema relacional con los mismos atributos simples y los subatributos hoja de los atributos compuestos.



¿A qué tabla se mapea?

biblioteca(nombre, *calle*, *número*, *ciudad*)

1.12.3. Regla CEF3

Un atributo multivalorado M simple de un CE E es representado por un esquema separado EM. EM tiene atributos correspondientes a la clave primaria de E y un atributo correspondiente al atributo multivalorado M. Todos los atributos de EM forman su clave primaria. Se pone una restricción de clave foránea desde EM que referencia a la clave primaria de E.



¿A qué tablas se mapea?

libro(título, ISBN, *editorial*, *edición*)
libro-autor(ISBN, autor)

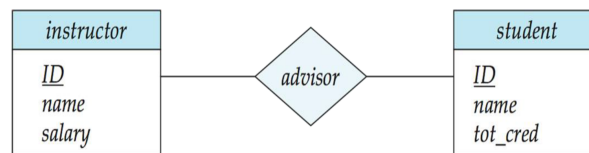
Aclaración: Cada valor del atributo multivalorado mapea a una tupla separada en la tabla del esquema EM.

Por ejemplo un libro con ISBN 970-17-0256-5 y autores con DNIs: 11122256 y 15345678 se mapea a dos tuplas: (970-17-0256-5, 11122256) y (970-17-0256-5, 15345678)

Decisión: Los atributos derivados no son explícitamente representados en el modelo de datos relacional. Se verá que si se los necesita una forma de computarlos es por medio de consultas.

1.12.4. Regla CR1

Un CR **varios a varios** es representado con un esquema con atributos para las claves primarias de los dos CE participantes y todos los atributos descriptivos del CR (que no son multivalorados). La clave primaria del esquema del CR es la unión de las claves primarias de los CE que participan en el CR. Para cada CE que participa en el CR se crea restricción de clave foránea que referencia clave primaria de CE.

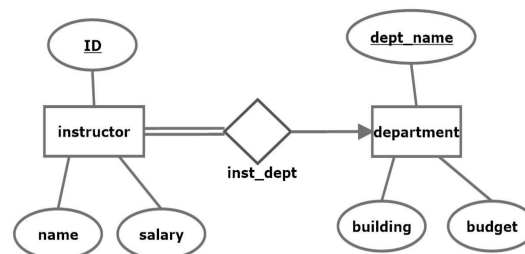


¿A qué esquema se mapea el CR?

$Advisor = (\underline{i-id}, \underline{s-id})$

1.12.5. Regla CR2

Un CR **varios a uno** o **uno a varios** que es total en el lado varios puede ser representado agregando atributos extra en el CE del lado varios, conteniendo la clave primaria del lado uno. La clave primaria del CR es la clave primaria del CE del lado varios. Se crea restricción de clave foránea de CR que referencia a clave primaria de CE de lado varios.



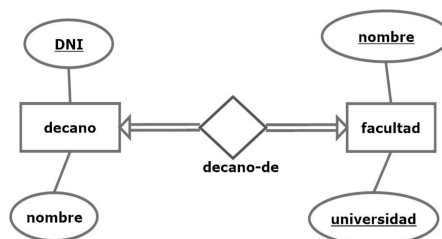
¿A qué tabla se mapea inst_dept?

$instructor(\underline{ID}, name, salary, dept_name)$

Si la participación es parcial en el lado varios, aplicar la regla anterior puede resultar en valores nulos. Esto sucede cuando a una entidad del CE del lado varios no le corresponde ninguna entidad del CE del lado uno.

1.12.6. Regla CR3

Un CR **uno a uno** puede ser mapeado agregando al esquema resultante de traducir uno de los CE participantes los atributos de la clave primaria del otro CE. La clave primaria de cualquier CE puede ser elegida como la clave primaria del CR. Se crea restricción de clave foránea de esquema relacional asociado al CR que referencia clave primaria de otro CE (el que no se tomo de base para hallar el esquema asociado al CR).



¿A qué tabla se mapea decano-de?

Las siguientes respuestas son igualmente válidas:

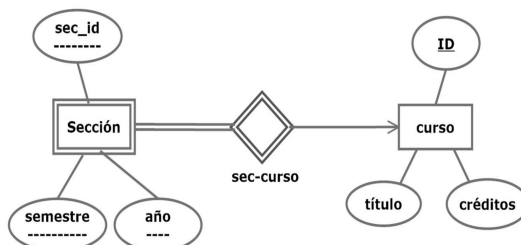
decano(DNI, nombre, nombreFacultad, universidad)

y

facultad (nombre, universidad, DNI)

1.12.7. Regla CED

Un CE débiles se mapea a una tabla que incluye columnas para la clave primaria del CE identificador más los atributos (no multi-valorados) del CE débiles (achatando jerarquías de atributos compuestos si es necesario). La clave primaria del CE identificador más el discriminador del CE débil forman la clave primaria del esquema relacional de la traducción. Para atributos de esquema de CE débil que provienen de CE identificadora se agrega restricción de clave foránea desde esquema de CE débil a CE identificador. El CR identificador no se mapea.



¿A qué tabla se mapea el CE débil?

sección(ID, sec_id, semestre, año)

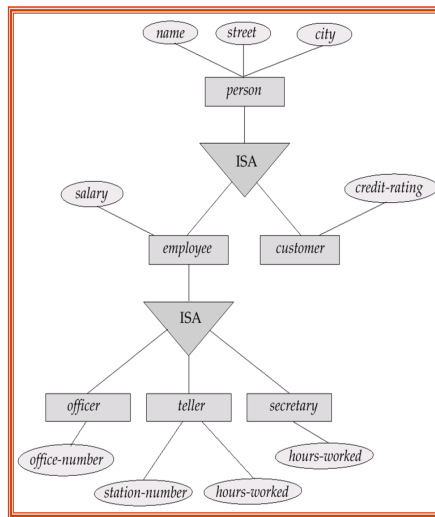
1.12.8. Reducción de generalización

Para traducir generalización iremos proponiendo y evaluando ideas de traducción. Al final decimos bajo qué circunstancias las ideas de traducción son reglas a usar.

Idea G1:

traducción de generalización a tablas.

- Formar una tabla para el CE de nivel más alto (la generalización)
- Formar una tabla para cada CE especialización que incluye la clave primaria del CE generalización y los atributos locales.



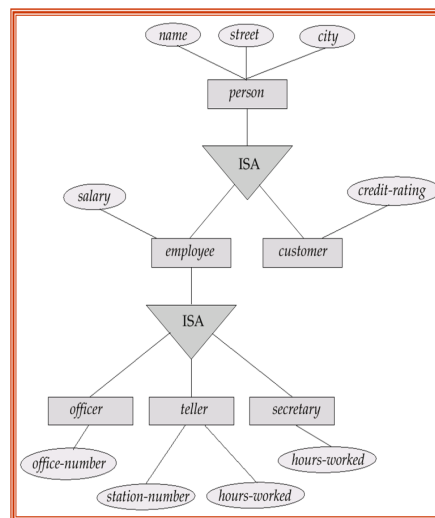
¿A qué tablas se mapea la generalización person?

person(name, street, city)
customer(name, credit-rating)
Employee(name, salary)

Evaluación de Idea G1: obtener toda la información de CE de especialización requiere acceder a dos tablas.

Idea G2:

formar una tabla para cada CE especialización con los atributos locales y heredados del CE generalización.



¿A qué tablas se mapea la generalización person usando idea G2?

person(name, street, city)
customer(name, street, city, credit-rating)
Employee(name, street, city, salary)

Evaluación de Idea G2: si dos o más especializaciones contienen la misma entidad, entonces los atributos que no son clave primaria del CE generalización son almacenados redundantemente.

Si la generalización es total con la traducción usando esta idea, la tabla del CE generalización no se requiere para almacenar información.

Idea G3:

si la generalización es total formar una tabla para cada CE especialización con los atributos locales y heredados; no formar tabla para la generalización. Esta idea es óptima para cuando la generalización es total y disjunta.

En resumen:

- Generalización total y disjunta: usar idea G3.
- Generalización no total y disjunta: usar idea G2.
- Generalización no disjunta: usar idea G1 (para evitar tener que tratar con redundancia de información).

Capítulo 2

Algebra Relacional