

Tema 7: Normalización de Modelos Relacionales

Introducción a la Ingeniería del Software y los Sistemas de Información I
Ingeniería Informática – Tecnologías Informáticas
Departamento de Lenguajes y Sistemas Informáticos



1. Anomalías de manipulación

2. Dependencias funcionales

3. Formas Normales

- La calidad de un modelo relacional depende, entre otros factores, de las **anomalías de manipulación** que presente.
- La forma de asegurar la calidad de un modelo relacional frente a las anomalías de manipulación es comprobar que está al menos en **tercera forma normal (3FN)**.

Anomalías de manipulación

- Supongamos una relación que contiene los datos de los inmuebles de una agencia de alquiler.
- Cada inmueble tiene un código, una dirección, un precio de alquiler, una lista de propietarios con el porcentaje de propiedad del inmueble, y el código, nombre y cargo del empleado que lo gestiona.

Inmuebles						
inmuebleId	dirección	precio	propietarios	empleadId	nombre	cargo
0010A	Avda. Reina mercedes, 15	600	P. González, 70% D. Páez, 30%	3	S. Armesto	Resp. Zona
2230A	Calle Tarifa, 15	500	E. Martos, 100%	3	S. Armesto	Resp. Zona
3387B	Los Bermejales, 8	700	R. Vidal, 50% P. González, 50%	5	N. Martín	Resp. Zona
7891A	Avda. de las Ciencias, 10	650	M. Gallego, 40% M. Sánchez, 60%	8	G. González	Comercial
0023B	Calle Telémaco, 14	800	R. Borrego, 70% J. Trajano, 30%	8	G. González	Comercial

- **Datos redundantes:** el nombre y el cargo de cada empleado se repita tantas veces como inmuebles gestione, malgastando espacio.
- **Riesgos de incoherencia:** la redundancia de datos implica el riesgo de que se vuelvan incoherentes si no se actualizan todas las ocurrencias a la vez.

Inmuebles						
inmuebleld	dirección	precio	propietarios	empleadold	nombre	cargo
0010A	Avda. Reina mercedes, 15	600	P. González, 70% D. Páez, 30%	3	S. Armesto	Resp. Zona
2230A	Calle Tarifa, 15	500	E. Martos, 100%	3	S. Armesto	Resp. Zona
3387B	Los Bermejales, 8	700	R. Vidal, 50% P. González, 50%	5	N. Martín	Resp. Zona
7891A	Avda. de las Ciencias, 10	650	M. Gallego, 40% M. Sánchez, 60%	8	G. González	Comercial
0023B	Calle Telémaco, 14	800	R. Borrego, 70% J. Trajano, 30%	8	G. González	Comercial

- Anomalías de inserción: hasta que un empleado no gestione un inmueble no se puede registrar en el sistema de información.
- Anomalías de actualización: si un empleado cambia de cargo hay que actualizarlo múltiples veces en lugar de hacerlo una sola vez.

Inmuebles						
inmuebleld	dirección	precio	propietarios	empleadold	nombre	cargo
0010A	Avda. Reina mercedes, 15	600	P. González, 70% D. Páez, 30%	3	S. Armesto	Resp. Zona
2230A	Calle Tarifa, 15	500	E. Martos, 100%	3	S. Armesto	Resp. Zona
3387B	Los Bermejales, 8	700	R. Vidal, 50% P. González, 50%	5	N. Martín	Resp. Zona
7891A	Avda. de las Ciencias, 10	650	M. Gallego, 40% M. Sánchez, 60%	8	G. González	Comercial
0023B	Calle Telémaco, 14	800	R. Borrego, 70% J. Trajano, 30%	8	G. González	Comercial

- Anomalías de eliminación: si un empleado deja de gestionar inmuebles, sus datos desaparecen del sistema de información.
- Problemas de consulta: ¿cómo se podrían conocer todos los inmuebles de un determinado propietario?

Inmuebles						
inmuebleId	dirección	precio	propietarios	empleadId	nombre	cargo
0010A	Avda. Reina mercedes, 15	600	P. González, 70% D. Páez, 30%	3	S. Armesto	Resp. Zona
2230A	Calle Tarifa, 15	500	E. Martos, 100%	3	S. Armesto	Resp. Zona
3387B	Los Bermejales, 8	700	R. Vidal, 50% P. González, 50%	5	N. Martín	Resp. Zona
7891A	Avda. de las Ciencias, 10	650	M. Gallego, 40% M. Sánchez, 60%	8	G. González	Comercial
0023B	Calle Telémaco, 14	800	R. Borrego, 70% J. Trajano, 30%	8	G. González	Comercial

- ¿Cuántos problemas de los anteriores se **evitan** con el nuevo modelo relacional de dos relaciones?
- ¿Qué modelo relacional es **mejor**?
- ¿Puede mejorarse **más**?

Inmuebles				
inmuebleld	dirección	precio	propietarios	empleadold
0010A	Avda. Reina mercedes, 15	600	P. González, 70% D. Páez, 30%	3
2230A	Calle Tarifa, 15	500	E. Martos, 100%	3
3387B	Los Bermejales, 8	700	R. Vidal, 50% P. González, 50%	5
7891A	Avda. de las Ciencias, 10	650	M. Gallego, 40% M. Sánchez, 60%	8
0023B	Calle Telémaco, 14	800	R. Borrego, 70% J. Trajano, 30%	8

Empleados		
empleadold	nombre	cargo
3	S. Armesto	Resp. Zona
5	N. Martín	Resp. Zona
8	G. González	Comercial

- Si **R** es una relación y **X** e **Y** son dos subconjuntos de los atributos de R, se dice que:

- X determina funcionalmente a Y
- Y depende funcionalmente de X

- Si y sólo si...

- **Siempre** que dos tuplas tienen los mismos valores de X, también tienen los mismos valores de Y.

$$\forall t_1, t_2 \in \text{extensión}(R) \cdot (t_1.X = t_2.X) \Rightarrow (t_1.Y = t_2.Y)$$

- Se denota como: $X \rightarrow Y$

- En otras palabras: **nunca** dos tuplas con los mismos valores de X pueden tener distintos valores de Y.

$$\nexists t_1, t_2 \in \text{extensión}(R) \cdot (t_1.X = t_2.X) \wedge (t_1.Y \neq t_2.Y)$$

¿Cómo se identifican las dependencias funcionales?

- Las dependencias funcionales **no pueden deducirse de los datos** de la extensión de una relación.
- Sólo podría descartarse su existencia si los datos de la extensión las contradijeran.
- Por lo tanto...
 - Las dependencias funcionales **dependen de la semántica** de los atributos de las relaciones en el modelo conceptual y, por extensión, en el **dominio del problema**.

Dependencias Funcionales

$\text{inmuebleld} \rightarrow \text{dirección}$

$\text{inmuebleld} \rightarrow \{\text{empleadold}, \text{nombre}, \text{cargo}\}$

$\text{inmuebleld} \rightarrow \{\text{dirección}, \text{precio}, \text{propietarios}\}$ $\{\text{inmuebleld}, \text{precio}\} \rightarrow \text{empleadold}$

$\text{empleadold} \rightarrow \{\text{nombre}, \text{cargo}\}$

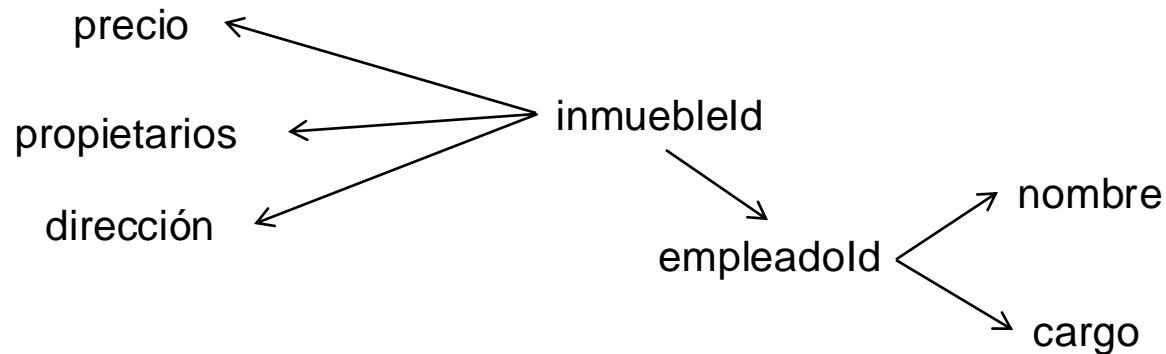
$\{\text{empleadold}, \text{nombre}\} \rightarrow \text{cargo}$

...

Inmuebles						
inmuebleld	dirección	precio	propietarios	empleadold	nombre	cargo
0010A	Avda. Reina mercedes, 15	600	P. González, 70% D. Páez, 30%	3	S. Armesto	Resp. Zona
2230A	Calle Tarifa, 15	500	E. Martos, 100%	3	S. Armesto	Resp. Zona
3387B	Los Bermejales, 8	700	R. Vidal, 50% P. González, 50%	5	N. Martín	Resp. Zona
7891A	Avda. de las Ciencias, 10	650	M. Gallego, 40% M. Sánchez, 60%	8	G. González	Comercial
0023B	Calle Telémaco, 14	800	R. Borrego, 70% J. Trajano, 30%	8	G. González	Comercial

Grafo de dependencias funcionales

- Forma gráfica de representar las dependencias funcionales de un modelo relacional.
- Los **nodos** son **atributos** o conjuntos de atributos.
- Los **arcos** son las **dependencias funcionales**.
- Normalmente sólo se representan dependencias funcionales que determinan a un solo atributo.



- Son **condiciones**, basadas en las dependencias funcionales, que debe cumplir un modelo relacional **para estar exento de anomalías de manipulación.**
- Originalmente, Codd propuso tres formas normales: **1FN, 2FN y 3FN.**
- Posteriormente, se han propuesto otras tres: Boyce-Codd FN, 4FN y 5FN.
- **Cada FN incluye a la anterior**, por lo que un modelo relacional en 3FN está también en 2FN y en 1FN.

- Una relación está en **1FN** si en cada tupla se le asigna a cada atributo **un solo valor** del dominio sobre el que está definido.
- Esto implica la **ausencia** de **grupos repetidos**.
- **Ejemplo 1FN:**
 - Pasar de un solo teléfono por cliente a varios.

Clientes		
clienteld	nombre	teléfono
1	Abel Abad	666111222
2	Braulio Brío	666222333
3	Carlos Cepa	666333444
..

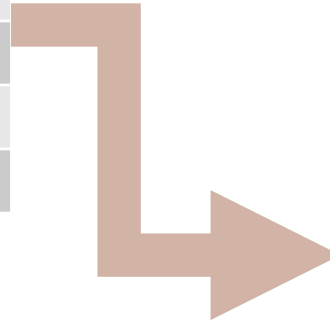


Clientes		
clienteld	nombre	teléfono
1	Abel Abad	666111222
2	Braulio Brío	666222333 666555666 954456789
3	Carlos Cepa	666333444 954123123
..

1FN FAIL

- Pasar de un teléfono por cliente a varios.*

Clientes		
clienteld	nombre	teléfono
1	Abel Abad	666111222
2	Braulio Brío	666222333
3	Carlos Cepa	666333444
..



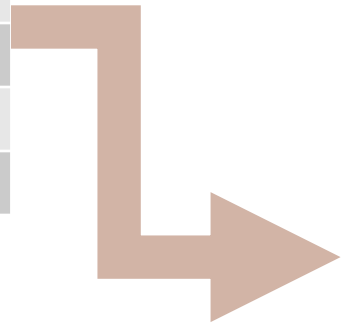
Clientes				
clienteld	nombre	teléfono1	teléfono2	teléfono3
1	Abel Abad	666111222	null	null
2	Braulio Brío	666222333	666555666	954456789
3	Carlos Cepa	666333444	954123123	null
..		

1FN FAIL

- Se desconoce cuántos teléfonos almacenar (“varios”)

- Pasar de un teléfono por cliente a varios.*

Clientes		
clienteld	nombre	teléfono
1	Abel Abad	666111222
2	Braulio Brío	666222333
3	Carlos Cepa	666333444
..



Clientes	
clienteld	nombre
1	Abel Abad
2	Braulio Brío
3	Carlos Cepa
..	...

Teléfonos	
clienteld	teléfono
1	666111222
2	666222333
2	666555666
2	954456789
3	666333444
3	954123123
..	...



- Relación de inmuebles para alquilar

Inmuebles						
ID_INMUEB	dirección	precio	propietarios	ID_EMP	nombre	cargo
0010A	Avda. Reina mercedes, 15	600	P. González, 70% D. Páez, 30%	3	S. Armesto	Resp. Zona
2230A	Calle Tarifa, 15	500	E. Martos, 100%	3	S. Armesto	Resp. Zona
3387B	Los Bermejales, 8	700	R. Vidal, 50% P. González, 50%	5	N. Martín	Resp. Zona

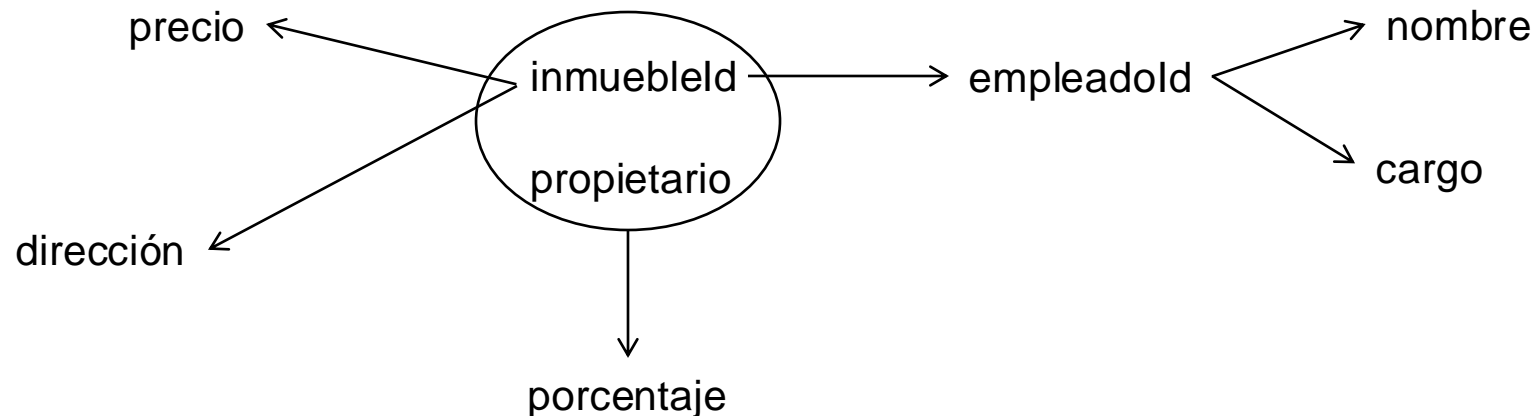


Inmuebles							
ID_INMUEB	dirección	precio	propietario	porcentaje	ID_EMP	nombre	cargo
0010A	Avda. Reina mercedes, 15	600	P. González	70%	3	S. Armesto	Resp. Zona
0010A	Avda. Reina mercedes, 15	600	D. Páez	30%	3	S. Armesto	Resp. Zona
2230A	Calle Tarifa, 15	500	E. Martos	100%	3	S. Armesto	Resp. Zona
3387B	Los Bermejales, 8	700	R. Vidal	50%	5	N. Martín	Resp. Zona
3387B	Los Bermejales, 8	700	P. González	50%	5	N. Martín	Resp. Zona

- Una relación está en **2FN** si está en 1FN y todos los **atributos no primos** son completamente dependientes de las claves candidatas de la relación.
- Los **atributos no primos** son los que no forman parte de ninguna clave candidata.
- Es decir, todos los atributos no clave deben depender completamente de las claves candidatas (no puede haber dependencia parcial).
- Normalmente una relación no está en 2FN porque está representando varias entidades y asociaciones a la vez.
- **Siempre** se puede transformar un modelo relacional a 2FN sin pérdidas de información ni dependencias.

Ejemplo 2FN:

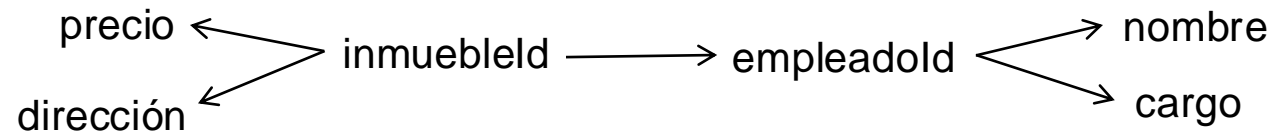
Inmuebles							
Primary Key: {inmuebleld, propietario}							
inmuebleld	dirección	precio	propietario	porcentaje	empleadold	nombre	cargo
0010A	Avda. Reina mercedes, 15	600	P. González	70%	3	S. Armesto	Resp. Zona
0010A	Avda. Reina mercedes, 15	600	D. Páez	30%	3	S. Armesto	Resp. Zona
2230A	Calle Tarifa, 15	500	E. Martos	100%	3	S. Armesto	Resp. Zona
3387B	Los Bermejales, 8	700	R. Vidal	50%	5	N. Martín	Resp. Zona
3387B	Los Bermejales, 8	700	P. González	50%	5	N. Martín	Resp. Zona



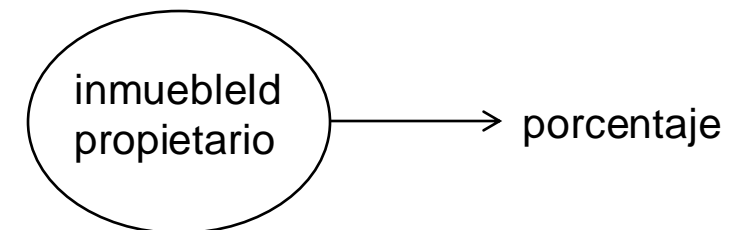
Ejemplo 2FN:

Inmuebles Primary Key: inmuebleld					
inmuebleld	dirección	precio	empleadold	nombre	cargo
0010A	Avda. Reina mercedes, 15	600	3	S. Armesto	Resp. Zona
2230A	Calle Tarifa, 15	500	3	S. Armesto	Resp. Zona
3387B	Los Bermejales, 8	700	5	N. Martín	Resp. Zona

2FN OK



Propietarios Primary Key: {inmuebleld, propietario} Foreign Key: inmuebleld / Inmuebles		
inmuebleld	propietario	porcentaje
0010A	P. González	70%
0010A	D. Páez	30%
2230A	E. Martos	100%
3387B	R. Vidal	50%
3387B	P. González	50%



2FN OK

Regla general para la 2FN

Si en la relación $R(\underline{K_1}, \underline{K_2}, X, Y)$ se tienen:

- los conjuntos de atributos primos: $\underline{K_1}$ y $\underline{K_2}$
- los conjuntos de atributos no primos: X e Y
- las dependencias funcionales: $\underline{K_1} \rightarrow X$ y $\{\underline{K_1}, \underline{K_2}\} \rightarrow Y$

Entonces:

- R no está en 2FN porque X no depende completamente de las claves candidatas, pero...
- La siguiente descomposición sí está en 2FN:
 - $R_1(\underline{K_1}, X)$ con $\underline{K_1} \rightarrow X$
 - $R_2(\underline{K_1}, \underline{K_2}, Y)$ con $\{\underline{K_1}, \underline{K_2}\} \rightarrow Y$

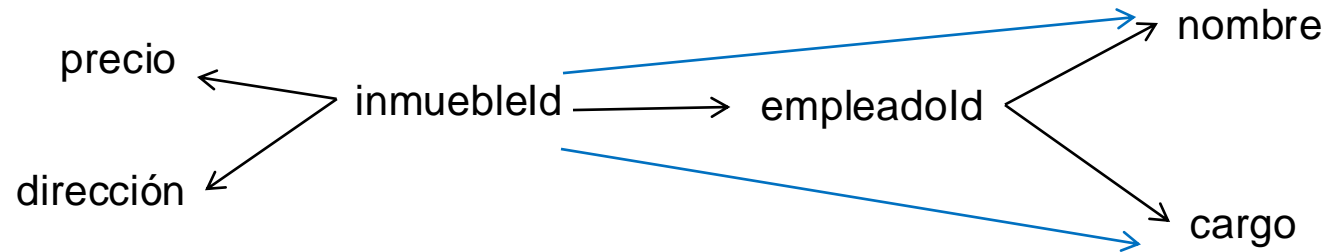
Tercera forma normal (3FN)

- Una relación está en **3FN** si está en 2FN y ningún atributo no primo depende **transitivamente** de ninguna clave candidata.
- Justificación de la 3FN
 - *Todos los atributos no primos deben representar un hecho sobre la clave, toda la clave y nada más que la clave.**
 - Normalmente una relación no está en 3FN porque está representando varias entidades asociadas a la vez.

* Fuente: https://es.wikipedia.org/wiki/Tercera_forma_normal

Ejemplo 3FN:

Inmuebles Primary Key: inmuebleId					
inmuebleId	dirección	precio	empleadId	nombre	cargo
0010A	Avda. Reina mercedes, 15	600	3	S. Armesto	Resp. Zona
2230A	Calle Tarifa, 15	500	3	S. Armesto	Resp. Zona
3387B	Los Bermejales, 8	700	5	N. Martín	Resp. Zona

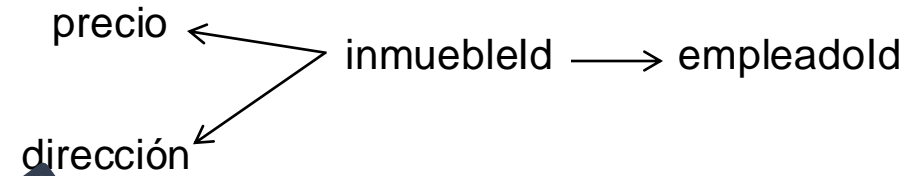


* Dependencia transitiva →

Ejemplo 3FN:

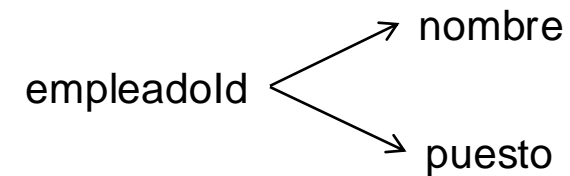
Inmuebles Primary Key: inmuebleId Foreign Key: empleadId / Empleados			
inmuebleId	dirección	precio	ID_EMP
0010A	Avda. Reina mercedes, 15	600	3
2230A	Calle Tarifa, 15	500	3
3387B	Los Bermejales, 8	700	5

3FN OK



Empleados Primary Key: empleadId		
empleadId	nombre	cargo
3	S. Armesto	Resp. Zona
3	S. Armesto	Resp. Zona
5	N. Martín	Resp. Zona

3FN OK



El fallo estaba en el modelo

Modelo A

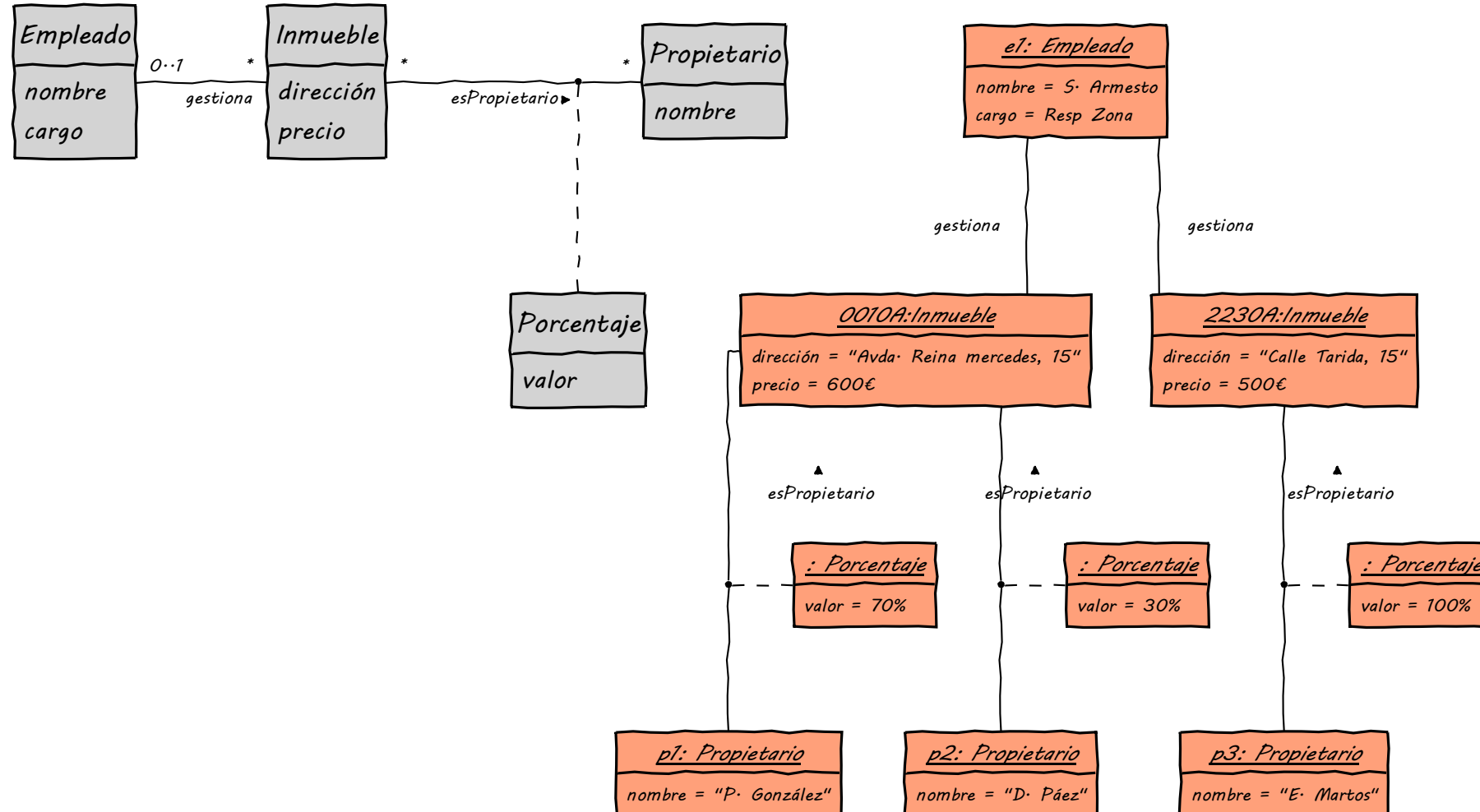
Inmueble
dirección
precio
propietarios
nombre
cargo

<u>0010A:Inmueble</u>
dirección = "Avda. Reina mercedes, 15"
precio = 600€
propietarios = {"P. González", 30%}, {"D. Páez", 30%}
nombre = S. Armesto
cargo = Resp. Zona

<u>2230A:Inmueble</u>
dirección = "Calle Tarida, 15"
precio = 500€
propietarios = {"E. MArtos", 100%}
nombre = S. Armesto
cargo = Resp. Zona

El fallo estaba en el modelo

Modelo B



Tema 7: Normalización de Modelos Relacionales

Introducción a la Ingeniería del Software y los Sistemas de Información I
Ingeniería Informática – Tecnologías Informáticas
Departamento de Lenguajes y Sistemas Informáticos

