

Modelización conceptual de datos. Modelo entidad relación.

1. MODELO DE DATOS.....	2
1.1. CONCEPTO.....	2
1.2. DEFINICIÓN FORMAL.....	2
1.2.1. <i>Componente estática</i>	2
1.2.2. <i>Componente dinámica</i>	3
2. MODELIZACIÓN CONCEPTUAL EN EL DISEÑO DE BASES DE DATOS.....	3
3. ANÁLISIS Y DISEÑO CONCEPTUAL.....	4
3.1. ANÁLISIS DE REQUISITOS.....	4
3.2. ETAPA DE CONCEPTUALIZACIÓN.....	5
4. MODELO ENTIDAD/INTERRELACIÓN.....	5
4.1. MODELO BÁSICO.....	5
4.1.1. <i>Entidad</i>	5
4.1.2. <i>Interrelación</i>	6
4.1.3. <i>Dominio y valor</i>	7
4.1.4. <i>Atributo</i>	7
4.2. MODELO EXTENDIDO.....	8
4.2.1. <i>Cardinalidad</i>	8
4.2.2. <i>Entidades e interrelaciones débiles</i>	8
4.2.3. <i>Entidades tipos y subtipos</i>	9
4.2.3.1. <i>Tipos de interrelaciones jerárquicas</i>	9

1. Modelo de datos.

1.1. Concepto.

Adoptando un punto de vista relacionado con el análisis de datos, se denomina modelo al instrumento que se aplica a una parcela del mundo real, al que siguiendo a Adoración de Miguel y Mario Piattini se va a denominar *universo del discurso (UD)*, para obtener una estructura de datos a la que se llamará esquema. Para evitar confusiones resulta necesario distinguir entre el modelo, que es el método o herramienta usado, y el esquema, que se obtiene como resultado de aplicar el método.

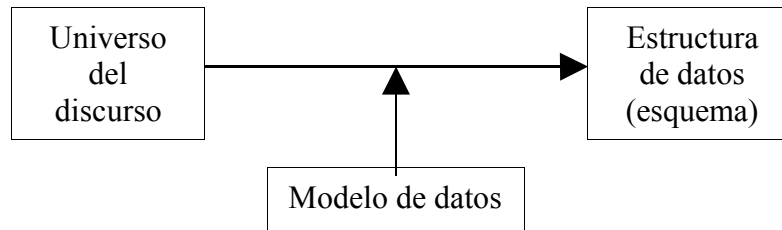


Figura 1. Modelo de datos

También hay que distinguir entre el mundo real y el universo del discurso, ya que este último no es la realidad en sí, sino la visión que de ella tiene el diseñador.

Hechas estas aclaraciones, es posible definir modelo de datos (MD) como un *conjunto de conceptos, reglas y convenciones que permiten describir los datos del universo del discurso* con el fin de permitir su representación en forma de datos en el sistema de información de la organización. El *objetivo* es aportar claridad y evitar las confusiones que surgen de intentar definir algo tan complejo como la estructura de una organización usando únicamente el lenguaje natural. De esta manera, el modelo contribuye a que el diseño sea lo más adaptado posible a la realidad que se pretende controlar.

1.2. Definición formal.

Desde un punto de vista de análisis temporal, un universo de discurso tiene dos tipos de características:

- Estáticas, denominadas estructuras.
- Dinámicas. Se trata de los datos o valores que se almacenan en las estructuras.

El modelo de datos puede entonces definirse formalmente como el par: MD (R,O)

donde R (reglas) representa la parte estática (estructura), y O (operaciones) la dinámica.

1.2.1. Componente estática.

La estática de un modelo de datos está compuesta por dos tipos de elementos. Por un lado están los **objetos recogidos por el modelo**. Entre los más comunes

- Entidades.
- Atributos o propiedades de la entidades.
- Dominios sobre los que se definirán los atributos.
- Interrelaciones, esto es, asociaciones entre entidades.

La representación de estos objetos depende del modelo de datos aunque la forma más usual es mediante grafos (entidad/relación).

Por otro lado están **las restricciones** que impiden la presencia de ciertos objetos o el que se establezcan determinadas asociaciones entre los objetos permitidos. Se clasifican en:

- Inherentes al modelo: impuestas por el modelo por su propia definición.
- De usuario; a través de las cuales se capta la semántica del universo del discurso que se ha de modelar.

Cuando se aplica la parte estática de un MD a un determinado UD se obtiene su esquema según dicho modelo. Los lenguajes asociados al modelo utilizado, que permiten representar la parte estática del UD se conocen como lenguajes de definición de datos (DDL Data Definition Languages)

1.2.2. Componente dinámica.

La componente dinámica del modelo consta de un conjunto de operaciones que se definen sobre la estructura del correspondiente modelo de datos. Formalmente es posible distinguir dos tipos de operaciones:

- Selección: seleccionar es localizar una o más ocurrencias de una entidad.
- Acción: que se realiza sobre las ocurrencias de entidad previamente localizadas mediante una operación de selección. Las acciones típicas son inserción, borrado y modificación.

Los lenguajes asociados al modelo utilizado, que permiten representar la parte dinámica del universo del discurso, se conocen como lenguajes de manipulación de datos (DML Data Manipulation Languages).

2. Modelización conceptual en el diseño de bases de datos.

Los modelos de datos son un eficaz instrumento en el diseño de bases de datos. Los niveles de abstracción de la arquitectura ANSI facilitan el diseño de una base de datos, al proporcionar nuevos instrumentos que ayudan a la estructuración, paso a paso, del mundo real hasta llegar a la base de datos física.

En el estado actual de la técnica es conveniente, en el diseño de bases de datos, distinguir la fase de modelado conceptual, que es la descripción del mundo real (empresa o administración) de acuerdo con un modelo altamente semántico e **independiente del SGBD** en el que posteriormente se vaya a hacer la implementación de la base de datos, y la fase de diseño lógico, en la cual se ha de obtener un esquema que responda a la estructura lógica específica (en general, relacional) del SGBD que se aplique en cada caso, por lo que dicho esquema está sometido a las restricciones que imponga el modelo del SGBD en concreto.

La figura 2 representa la forma de llegar desde la parcela del mundo real que se está analizando a la base de datos física. En un primer paso, con la ayuda del modelo conceptual, se obtiene el *esquema conceptual*, aunque a veces no se formaliza este paso y el analista, sin una metodología precisa, hace una abstracción del mundo real, que es lo que hemos llamado estructura percibida (lo hemos representado con líneas de puntos para indicar que se trata de un camino alternativo, aunque en nuestra opinión no aconsejable). A continuación, aplicando al esquema conceptual las reglas del modelo de datos propio del SGBD que se va a utilizar, se obtiene el *esquema lógico* (también llamado *esquema de base de datos*); de éste se pasa al *esquema interno*, donde el objetivo es conseguir la

máxima eficiencia de cara a la máquina y al problema específico. Por último, se implementa la base de datos física en los soportes secundarios. La estructura física se ha de *rellenar* con los valores (ocurrencias) que se obtienen por observación de los sucesos del mundo real.

En el enfoque que acabamos de presentar nos estamos moviendo en tres mundos distintos: el de la realidad, el de las ideas y el de los datos; mundos que se confunden a menudo, de forma que todos pasamos a veces de uno a otro *sin advertencia previa*.

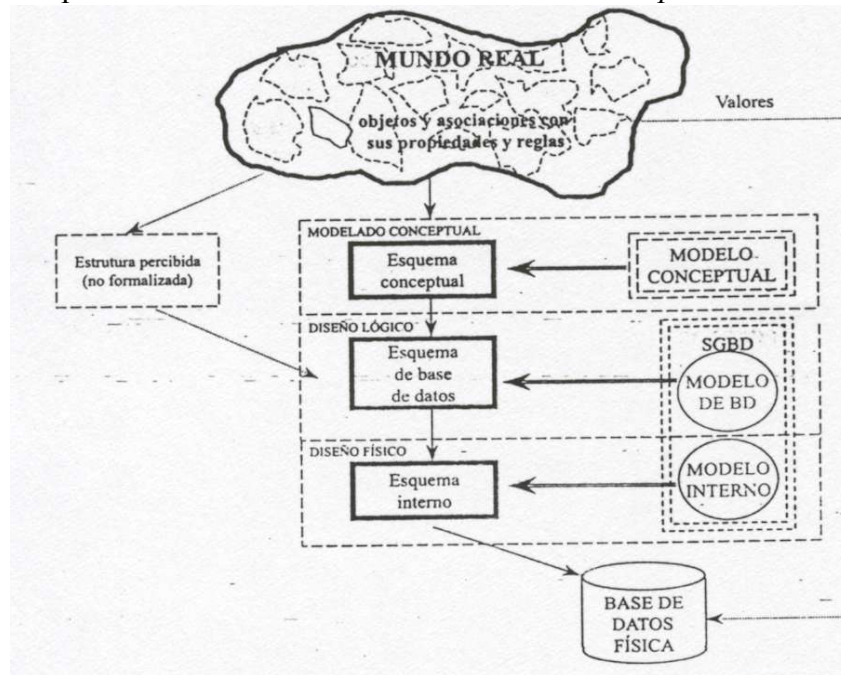


Figura 2. Detalle de la transformación del mundo real a la BD física

En el primero de estos dominios, el mundo real (al que también llamamos Universo del Discurso), existen objetos y asociaciones entre ellos; ambos tienen propiedades y hay reglas que imponen ciertas limitaciones. Será necesaria una abstracción de este mundo real por parte del diseñador de la base de datos que lo observa a fin de obtener un esquema conceptual, que será tanto más perfecto cuanto más se asemeje al mundo real. El modelo entidad/interrelación pertenece a este campo y recoge las entidades y las interrelaciones así como las restricciones semánticas.

Las herramientas CASE proporcionan una importante ayuda en el diseño de bases de datos, al disponer de modelos de datos semánticos (en general basados en el modelo E/R, con algunas extensiones) que facilitan el diseño conceptual y realizan la transformación al modelo relacional propio de los productos comerciales más extendidos.

3. Análisis y diseño conceptual.

El modelo conceptual constituye la primera fase de diseño y puede subdividirse en dos etapas claramente diferenciadas: análisis de requisitos y conceptualización.

3.1. Análisis de requisitos.

Es la etapa de percepción, identificación y descripción de los fenómenos del mundo real a analizar, es decir, se ha de responder a la pregunta ¿qué representar?

En esta etapa se procede a recoger la información relevante del universo del discurso a modelizar. Para ello se realizarán entrevistas a los usuarios de los diferentes niveles de organización. Se llevará a cabo un estudio de las llamadas reglas de gestión que recogerán las restricciones a contemplar. Por ejemplo: cada pedido tiene un único cliente, un pedido puede incluir uno o más productos, ... En esta etapa se recopilarán todos los documentos usados. Al final obtendremos un esquema descriptivo de la realidad, normalmente en lenguaje natural.

3.2. Etapa de conceptualización.

En ella se transforma el esquema descriptivo refinándose y estructurándose adecuadamente. Esta etapa responde a la pregunta ¿cómo representar?

En esta etapa obtendremos el esquema conceptual. Para representar este esquema se habrá de buscar una representación normalizada que se apoye en un modelo de datos que cumpla determinadas propiedades: coherencia, plenitud, no redundancia, simplicidad, ...

Para representar el esquema conceptual se utiliza el modelo entidad/interrelación, además de un diccionario de datos para describir el contenido de los objetos y una serie de fichas y plantillas.

4. Modelo Entidad/Interrelación.

El modelo entidad/interrelación (E/R) es, sin género de dudas, el modelo de datos conceptual más extendido en las metodologías de diseño de bases de datos y en las herramientas CASE. Fue propuesto por Peter Chen en 1976. El modelo E/R ha tenido una gran difusión y ha despertado un enorme interés entre los estudiosos de las bases de datos. Gracias a ello ha recibido las aportaciones de muchos autores, por lo que realmente no se puede considerar que exista un único modelo E/R, sino más bien una familia de modelos.

4.1. Modelo básico.

4.1.1. Entidad.

Se puede definir entidad como aquel objeto del universo del discurso sobre el que se desea almacenar información en la base de datos. Por ejemplo, una persona, lugar, cosa, concepto o suceso de interés para la empresa. Nosotros denominaremos a la estructura genérica en su sentido abstracto, tipo de entidad, mientras que entidad es cada una de las ocurrencias o instancias de ese tipo de entidad.

Para saber si una entidad pertenece a un tipo de entidad, analizaremos si cumple el predicado asociado al correspondiente tipo de entidad. Por ejemplo, el tipo de entidad LIBRO cuyo predicado asociado es “Documento editado que puede ser leído”, tiene como una ocurrencia “Concepción y Diseño de bases de datos” que pertenece a ella, ya que cumple dicho predicado.

La representación gráfica de un tipo de entidad es un rectángulo etiquetado con el nombre del tipo (normalmente un sustantivo).



Figura 3. Entidad LIBRO.

4.1.2. Interrelación.

Se entiende por interrelación aquella asociación o correspondencia existente entre entidades. Denominaremos tipo de interrelación a la estructura genérica del conjunto de interrelaciones existentes entre dos o mas tipos de entidad.

Representaremos el tipo de interrelación mediante un rombo etiquetado con el nombre de la interrelación, unido mediante arcos a los tipos de entidad que asocia (existen otras notaciones como la de Martín).

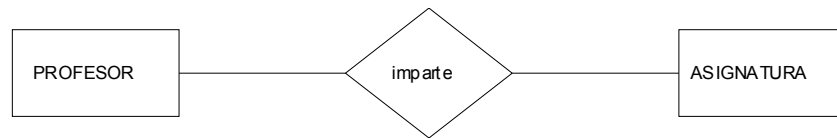


Figura 4. Interrelación.

Entre dos tipos de entidad puede existir más de un tipo de interrelación como puede observarse en el siguiente ejemplo:

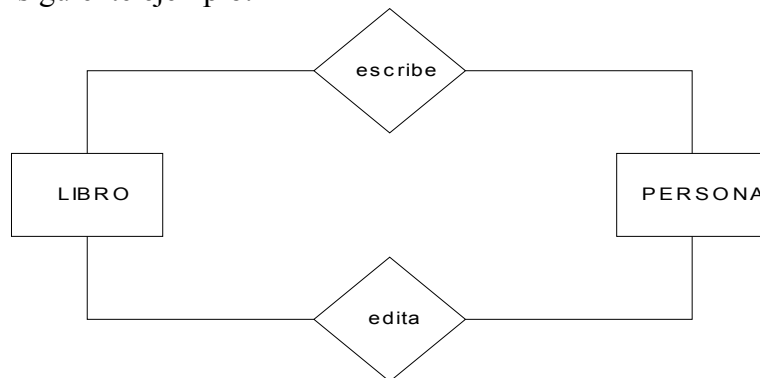
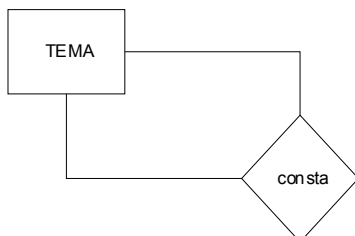


Figura 5. Dos interrelaciones distintas.

En un tipo de interrelación existen las siguientes elementos:

- **Nombre:** que lo distingue de los demás objetos. Debe ser único y significativo, normalmente un verbo.
- **Grado:** es el número de tipos de entidad que participan en un tipo de interrelación. Existen diversas posibilidades: reflexiva (una entidad se relaciona consigo misma), binaria, es decir, entre dos tipos de entidad (grado 2), n-aria, entre n tipos de entidad (grado n).



•

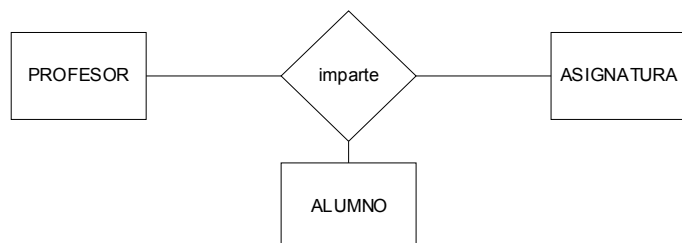


Figura6. Relación reflexiva monaria (grado 1)

Figura7. Relación ternaria (grado 3)

- **Tipo de correspondencia:** representa el número máximo de ocurrencias de cada tipo de entidad que pueden intervenir en el conjunto de ocurrencias del tipo de interrelación. Para representar gráficamente los casos posibles se pone una etiqueta con los valores 1:1, 1:N, N:M, según corresponda. Así por ejemplo, 1:N, significa que tomando todas las posibles ocurrencias de la interrelación, de uno de los tipos de entidad sólo aparecerá una ocurrencia que estará asociada con varias (máximo N no especificado) ocurrencias del otro tipo.
- **Papel (rol):** es la función que cada uno de los tipos de entidad realiza en el tipo de interrelación. El rol se indica dando un nombre al arco que une cada tipo de entidad con el tipo de interrelación.

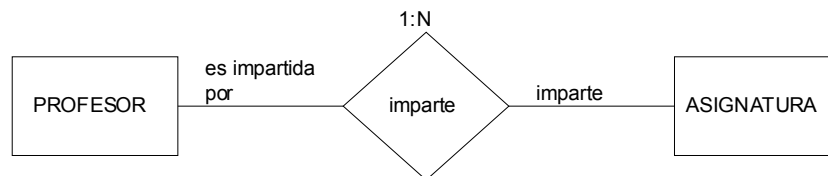


Figura 8. Relación binaria con tipo de correspondencia y rol

El significado es el siguiente: un profesor puede impartir varias asignaturas, pero una asignatura solo es impartida por un único profesor.

4.1.3. Dominio y valor.

El conjunto de posibles valores que puede tomar una propiedad dada de un tipo de entidad o interrelación se denomina dominio. Por ejemplo la edad de una persona puede estar comprendida entre [0 .. 150] años. Por lo que una edad de 15 años es válida.

4.1.4. Atributo.

Recibe el nombre de atributo cada una de las propiedades o características que tiene un tipo de entidad o un tipo de interrelación. Un atributo puede tomar valores de uno o más dominios.

Existen varias formas de mostrar los atributos en un diagrama E/R. La más extendida es la siguiente:

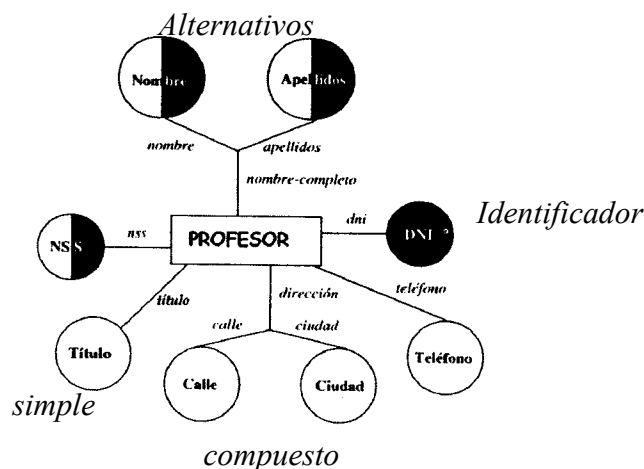


Figura 9. Representación de los distintos atributos.

Hay una serie de tipos de atributo que tienen un significado especial:

- **Atributo identificador candidato (AIC):** atributo o conjunto de atributos que permite identificar unívocamente cada ocurrencia de la entidad tipo. Ha de ser mínimo, es decir, si eliminamos cualquiera de los atributos individuales que lo componen la resultante no ha de ser, a su vez, AIC.
- **Atributo identificador principal (AIP):** aun atributo identificador candidato que se elige como identificador principal del tipo de entidad.
- **Atributo identificador alternativo (AIA):** si hay más de un AIC todos los que no sean AIP serán AIA.
- **Atributo compuesto:** se trata de un atributo formado a partir de varios dominios (o varios atributos simples).

Las interrelaciones también pueden tener atributos. Aparecen sobre todo cuando se necesita contemplar los requerimientos temporales.

4.2. Modelo extendido.

Diversos autores han propuesto extensiones al modelo E/R definido por Chen, dando lugar a lo que es posible denominar modelo extendido EE/R.

4.2.1. Cardinalidad.

Se define cardinalidad como el número máximo y mínimo de ocurrencias de un tipo de entidad que pueden estar interrelacionadas con una ocurrencia del otro, u otros tipos de entidad que participan en el tipo de interrelación. Su representación gráfica es una de las etiquetas (0,1), (1,1), (0,n) o (1,n), según corresponda.

La cardinalidad máxima coincide con el tipo de correspondencia definido por Chen.

Una cardinalidad mínima de 0, está indicando una interrelación opcional. Dada una ocurrencia de una de las entidades tipo que participa en dicha interrelación, si dicha entidad tipo tiene 0 de cardinalidad mínima, la citada ocurrencia podrá, a su vez, estar o no relacionada con otra u otras ocurrencias de las otras entidades que participan en la interrelación.

Una cardinalidad mínima 1 está indicando una interrelación obligatoria.

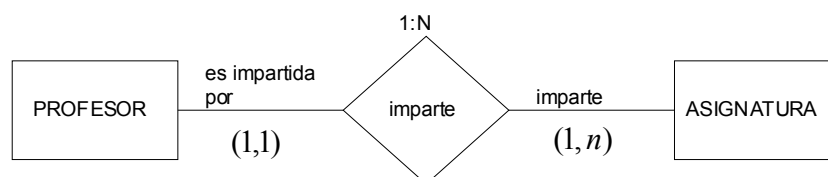


Figura 10. Interrelación con cardinalidades.

4.2.2. Entidades e interrelaciones débiles.

Los tipos de entidad se clasifican en:

- **Regulares o fuertes:** son aquellas cuyas ocurrencias no dependen, para existir, de la presencia de ocurrencias de ninguna otra entidad.
- **Débiles:** son aquellas cuyas ocurrencias sólo pueden aparecer cuando existen ocurrencias de una entidad regular de la que dependen.

- *Dependencia en existencia*: es inherente a la condición de debilidad de una entidad, es decir, la ocurrencia de la entidad débil no puede aparecer si no existe la correspondiente ocurrencia de la entidad fuerte con la que está relacionada.
- *Dependencia en identificación*: incluye la anterior pero además debe cumplirse que en el AIP del tipo de entidad débil participa el AIP de la entidad regular.

Para distinguir las interrelaciones débiles algunos autores proponen etiquetar el rombo con ID (identificación) y EX (existencia). Para diferenciar las entidades débiles de las regulares se incluye un doble rectángulo.

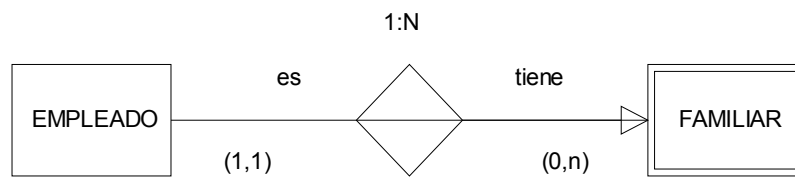


Figura 11. Entidad débil por existencia.

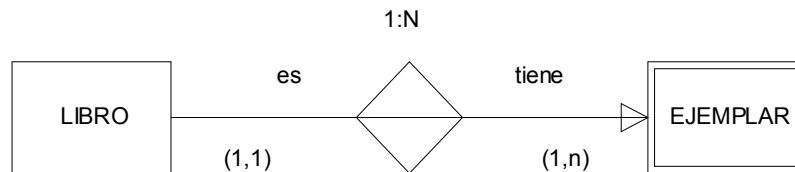


Figura 12. Entidad débil por identificación.

En el primer ejemplo, si desaparece la ocurrencia de empleado desaparecerán las ocurrencias de los familiares de ese empleado. No tiene sentido tener almacenados los familiares si el empleado ya no está en la empresa.

En el segundo caso, cada ejemplar (1,2, ... , n), solo tendrá sentido si existe el libro al que pertenece (existencia) y además solo podrá ser identificado a través de ese libro (identificación).

4.2.3. Entidades tipos y subtipos.

El modelo EE/R permite representar las relaciones jerárquicas existentes entre los tipos de entidad de los problemas del mundo real. La **generalización** es una abstracción que identifica una relación jerárquica. Nos indica que un tipo de entidad es un subtipo de otro tipo de entidad a un nivel de abstracción mayor, o en el sentido contrario, que un tipo de entidad es supertipo de otras de menor nivel de abstracción.

Existe una mutua dependencia entre las entidades subtipo y sus supertipos. Así, dada una ocurrencia de un supertipo, como máximo se relacionará con una ocurrencia de cada uno de los subtipos que le correspondan. A su vez, una ocurrencia de una entidad subtipo siempre estará relacionada con una, y solo una, entidad supertipo.

4.2.3.1. Tipos de interrelaciones jerárquicas.

Es posible diferenciar varias clases de generalización/especialización:

- Con o sin solapamiento: si en una ocurrencia de la interrelación tipo pueden aparecer ocurrencias de más de una entidad subtipo, además de la ocurrencia de la entidad supertipo, diremos que hay solapamiento. No existirá solapamiento cuando sólo pueda aparecer una ocurrencia de una entidad subtipo, además de la ocurrencia de la supertipo. A las interrelaciones sin solapamiento se las denomina **exclusivas**.
- Total o parcial: la generalización será total cuando una ocurrencia de la entidad supertipo esté asociada necesariamente a al menos una ocurrencia de una de las entidades subtipo. Habrá generalización parcial cuando puedan existir ocurrencias de la supertipo que aparezcan solas sin relacionarse con ninguna ocurrencia de alguna de las subtipo.

Estas categorías pueden mezclarse.

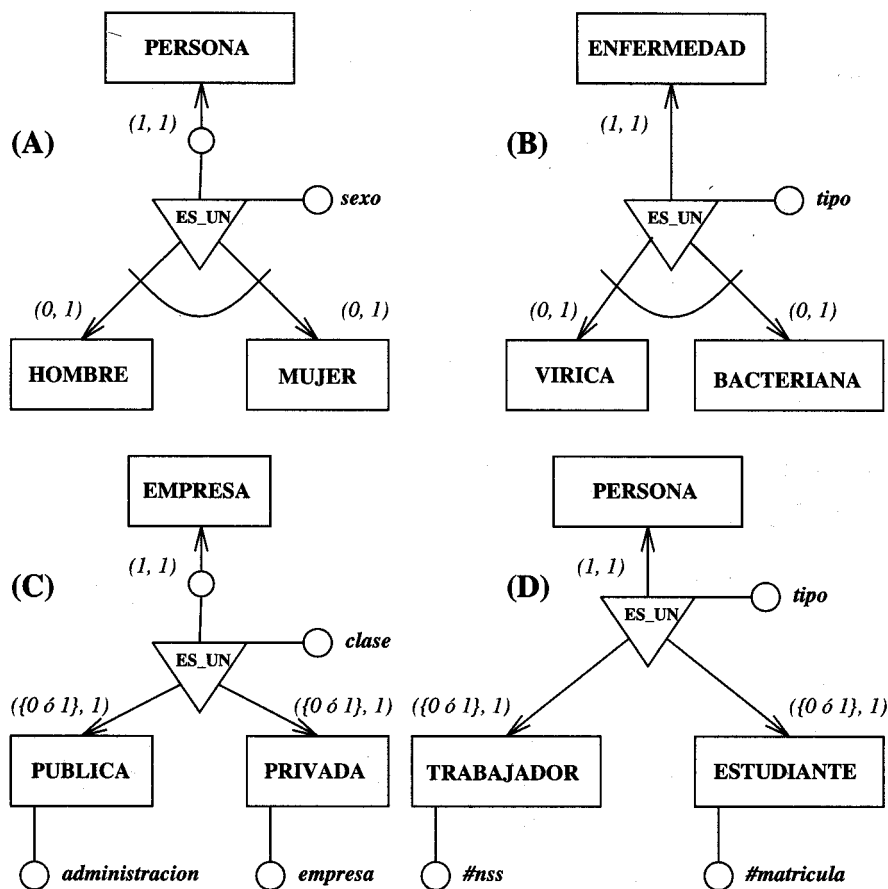


Figura 13. Tipos de interrelaciones jerárquicas.

- A) Sin solapamiento total.
- B) Sin solapamiento parcial.
- C) Con solapamiento total.
- D) Sin solapamiento parcial.