

Tema 2: Diseño conceptual de bases de datos.

2.1. Introducción.

El modelo de datos más empleado en la actualidad para llevar a cabo el diseño o modelado conceptual de bases de datos es el modelo Entidad-Interrelación, más conocido como modelo Entidad-Relación o incluso modelo E-R. Este modelo fue propuesto por Peter Chen en 1976. Posteriormente otros autores realizaron aportaciones al modelo básico propuesto por Chen, dando lugar a lo que se denomina el modelo extendido Entidad-Relación. En primer lugar se expondrá el modelo básico propuesto por Chen, dejando para un apartado posterior las extensiones propuestas por otros autores.

2.2. El modelo Entidad-Relación básico.

Los elementos básicos del modelo son: las entidades, los dominios y atributos y las relaciones o interrelaciones.

2.2.1. Entidad.

Podemos definir una entidad como cualquier objeto sobre el que se desea almacenar información en la base de datos. Por ejemplo, si queremos desarrollar una base de datos para gestionar la información de una biblioteca, podrían ser entidades los libros, los autores, las editoriales, etc.

Se llama ocurrencia de una entidad a cada una de las realizaciones concretas de esa entidad en el mundo real o a cada instancia de esa entidad. Así, por ejemplo, una ocurrencia de la entidad *Libro* podría ser el libro titulado “Bases de datos relacionales y modelado de datos”.

Una entidad se representa mediante un rectángulo en el interior del cual se coloca el nombre de la entidad en cuestión. Por ejemplo, la entidad *Libro* se representaría así:

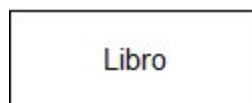


Figura 1: Representación gráfica de una entidad.

2.2.2. Relación.

Una relación o interrelación se puede definir como una asociación o correspondencia entre entidades. Así, si tenemos dos entidades *Empleado* y *Departamento*, se podría establecer una relación entre ellas llamada *trabaja* para reflejar qué empleados trabajan en cada departamento.

Una relación se representa mediante un rombo con el nombre de la relación en su interior y desde el que salen líneas que lo unen a las entidades participantes en la relación. Por ejemplo, la relación comentada en el párrafo anterior se puede representar así:

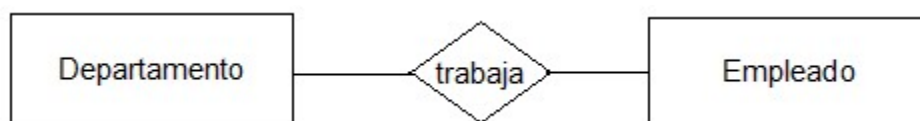


Figura 2: Representación gráfica de una relación

Toda relación tiene las siguientes características:

- Nombre: Toda relación debe tener un nombre único en el esquema E-R.
- Grado: Hace referencia al número de entidades que participan en una relación. En función del grado las relaciones se pueden clasificar en:
 - Relaciones binarias o de grado 2: Son aquellas que relacionan a dos entidades. Así, la relación de la figura 2 es una relación binaria.
 - Relaciones reflexivas o de grado 1: Son aquellas que relacionan una entidad consigo misma. Así, la relación de la figura 3 es una relación reflexiva.



Figura 3: Relación reflexiva.

- Relaciones ternarias, cuaternarias,... o de grado 3, 4,...: Son aquellas que relacionan tres, cuatro,... entidades respectivamente. En la figura 4 se representa una relación ternaria.

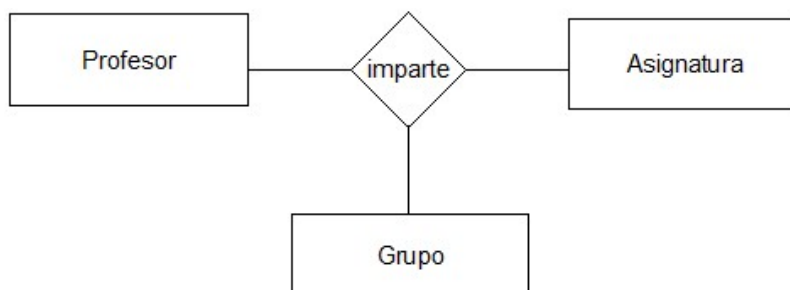


Figura 4: Relación ternaria.

Las relaciones más habituales son las binarias, y a continuación las reflexivas. Sin embargo, las relaciones de grado superior a 2 son bastante escasas.

- Tipo de correspondencia: Hace referencia al número máximo de ocurrencias de una entidad que pueden estar asociadas con una ocurrencia de la otra entidad participante en la relación. El tipo de correspondencia se indica al lado del rombo que representa la relación, y para las relaciones reflexivas y binarias puede ser:
 - 1:1: Se da cuando cada ocurrencia de una entidad solo puede estar asociada como máximo con una ocurrencia de la otra entidad. Ejemplos: Un director dirige un instituto y un instituto solo tiene un director. Una persona está casada con una sola persona.

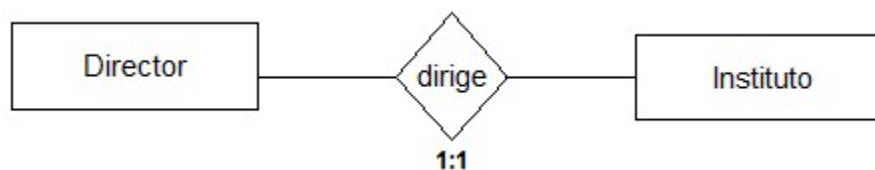


Figura 5: Relación 1:1 binaria.



Figura 6: Relación 1:1 reflexiva.

- 1:N: Se da cuando una ocurrencia de una entidad puede estar asociada con varias ocurrencias de la otra entidad, mientras que una ocurrencia de la otra entidad solo puede estar asociada con una ocurrencia de la primera. Ejemplos: En un departamento trabajan varios empleados, pero un empleado trabaja en un solo departamento. Un empleado puede ser jefe de varios empleados, pero un empleado tiene un solo jefe.

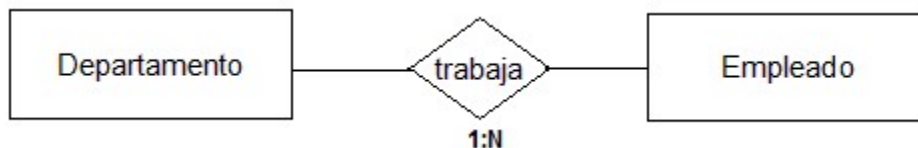


Figura 7: Relación 1:N binaria.

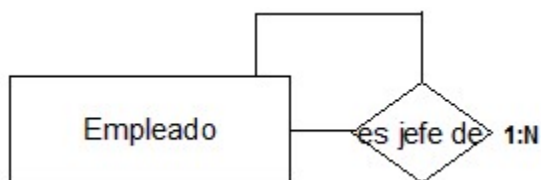


Figura 8: Relación 1:N reflexiva.

- **N:M:** Se da cuando una ocurrencia de una entidad puede estar asociada con varias ocurrencias de la otra entidad y cada ocurrencia de la otra entidad también puede estar asociada con varias ocurrencias de la primera. Ejemplos: En un pedido se incluyen varios artículos y un artículo puede ser solicitado en varios pedidos. Una pieza se puede descomponer en varias piezas y, a su vez, una pieza puede aparecer como componente de varias piezas.

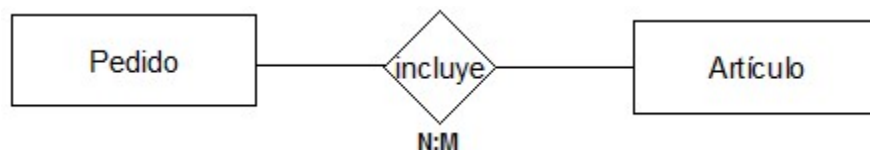


Figura 9: Relación N:M binaria.

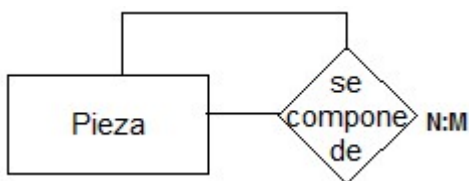


Figura 10: Relación N:M reflexiva.

2.2.3. Atributo y dominio.

Podemos definir atributo como cada una de las características o propiedades de una entidad o de una relación. Así, por ejemplo, la entidad *Libro* puede tener los atributos código, ISBN, título, número de páginas, etc.

El dominio de un atributo se puede definir como el conjunto de valores que puede tomar ese atributo. Así, por ejemplo, el dominio para el ISBN podría ser una cadena de 10 caracteres.

Los dominios se suelen representar mediante un óvalo o círculo con el nombre del mismo en su interior y el nombre del atributo en cuestión sobre la línea que une el óvalo o círculo con el rectángulo correspondiente a la entidad. No obstante, dado que esta representación haría que los esquemas E-R fuesen excesivamente densos y poco comprensibles, lo que se suele hacer es colocar en el interior del óvalo o círculo el nombre del atributo y omitir el nombre del dominio. En otros casos, se dibuja un pequeño círculo unido mediante una línea al rectángulo de la entidad, al lado del cual se escribe el nombre del atributo.

Podemos hablar de los siguientes tipos de atributos:

- Atributo identificador candidato (AIC): Es aquel atributo o conjunto de atributos que permite identificar unívocamente cada ocurrencia de la entidad correspondiente. Por ejemplo, para la entidad *Libro* podría ser atributo identificador candidato el ISBN porque no existen dos libros con el mismo ISBN y, por tanto, el ISBN sirve para identificar sin equivocación posible a cada libro.
- Atributo identificador principal (AIP): Es aquel atributo identificador candidato seleccionado para identificar a cada ocurrencia de la entidad. Si solo hay un AIC, ese es el único posible AIP. En nuestro ejemplo, el ISBN sería el AIP de la entidad *Libro*.
- Atributo identificador alternativo (AIA): Es aquel atributo identificador candidato no elegido como atributo identificador principal.

Existen diferentes maneras de representar en un esquema E-R los atributos identificadores principales y alternativos:

- Cuando se representan los nombres de los atributos en el interior de un óvalo o círculo, los nombres de los AIP se subrayan con trazado continuo y los nombres de los AIA se subrayan con trazado discontinuo.
- Cuando se usa un pequeño círculo para cada atributo, el círculo se representa relleno de color negro para los AIP, y la mitad del círculo de color negro para los AIA.

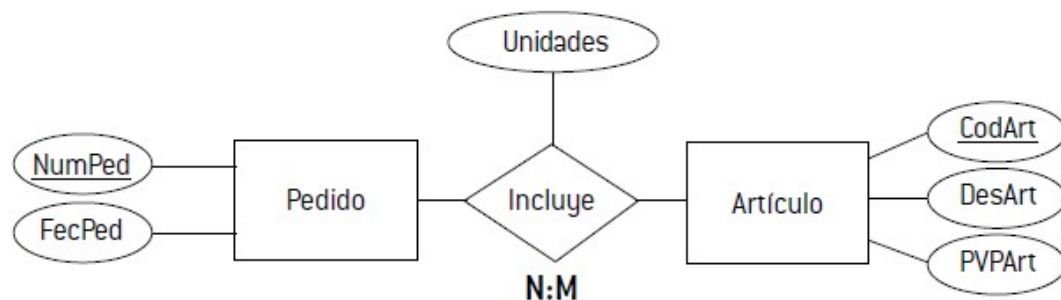


Figura 11. Representación de atributos

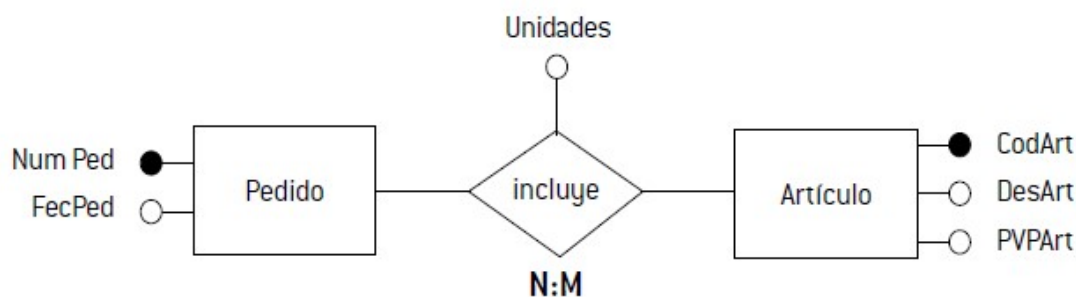


Figura 12. Representación de atributos alternativa

2.3. El modelo Entidad-Relación extendido.

Diversos autores han propuesto extensiones interesantes al modelo E-R básico propuesto por Peter Chen, dando lugar al modelo E-R extendido. Se estudian en este apartado las extensiones de mayor relevancia.

2.3.1. Cardinalidad de las relaciones.

Las cardinalidades mínima y máxima de las entidades intervinientes en una relación se definen como el número mínimo y máximo de ocurrencias de una entidad que pueden estar relacionadas con una ocurrencia de la otra entidad. Las cardinalidades se representan por medio de las etiquetas (0,1), (1,1), (0,n) o (1,n) sobre la línea que une a cada entidad con el rombo que representa la relación. Cuando la cardinalidad es (0,n) o (1,n) se suele dibujar una punta de flecha sobre la línea que une a la relación con la entidad correspondiente. Además, cuando se indican cardinalidades, se pueden omitir los tipos de correspondencia, como se va a hacer a partir de ahora.

En el ejemplo de la siguiente figura, el (1,n) al lado de la entidad *Artículo* indica que un pedido incluye como mínimo un artículo (1) y como máximo muchos (n), mientras que el (0,n) al lado de la entidad *Pedido* significa que un artículo puede no estar incluido en ningún pedido (0) o puede estar incluido en varios (n).

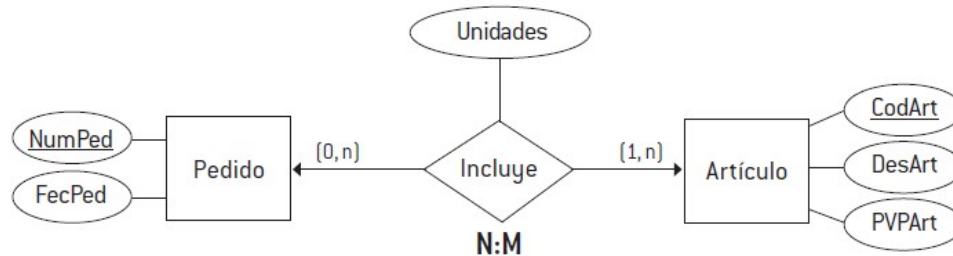


Figura 13: Representación de cardinalidades en una relación binaria N:M.

En el siguiente ejemplo, en el que hay una relación reflexiva, el (0,1) quiere decir que un empleado tiene 0 o un jefe. No tendrá ninguno si es el “jefe supremo de la empresa”, es decir, si no tiene a nadie por encima en la jerarquía de empleados de la empresa; uno, en cualquier otro caso. Por su parte, el (0,n) quiere decir que un empleado puede tener 0 o muchos subordinados. No tendrá ningún subordinado si no es jefe en la empresa, y si lo es, puede tener muchos subordinados.

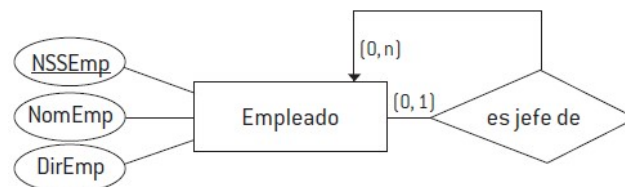


Figura 14: Representación de cardinalidades en una relación reflexiva 1:N.

En el siguiente ejemplo, un (0,n) significa que una pieza puede no descomponerse en otras o bien descomponerse en varias. No se descompondrá en ninguna si no es posible dividirla en piezas más pequeñas, es decir, si es indivisible. El otro (0,n) significa que una pieza puede ser componente o subordinada, como mínimo de ninguna (si no forma parte de ninguna otra pieza superior a ella) o, como máximo, de varias, en el caso de que esta pieza sea componente de varias piezas.

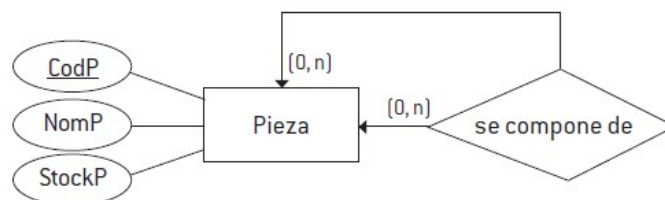


Figura 15: Representación de cardinalidades en una relación reflexiva N:M.

Para el caso concreto de las relaciones con grado mayor que 2 (relaciones ternarias, cuaternarias, etc.), las cardinalidades se obtienen de manera diferente. Consideremos la relación ternaria de la figura 16 entre las entidades *Suministrador*, *Proyecto* y *Componente*. Esta relación nos indica los componentes que proporciona cada suministrador para cada proyecto que lleva a cabo una empresa. El cálculo de las cardinalidades se lleva a cabo considerando una ocurrencia de cada par de entidades y determinando con cuántas ocurrencias de la otra entidad puede estar relacionado ese par como mínimo y como máximo. Así, en la figura 16:

- El (1,n) al lado de *Suministrador* quiere decir que para un proyecto en concreto un componente puede ser proporcionado como mínimo por un suministrador y como máximo por muchos.
- El (0,n) al lado de *Proyecto* indica que un suministrador un componente determinado puede que no lo suministre a ningún proyecto, a uno o incluso a muchos.
- El (0,n) al lado de *Componente* indica que un suministrador a un proyecto puede no proporcionarle ningún componente o muchos.

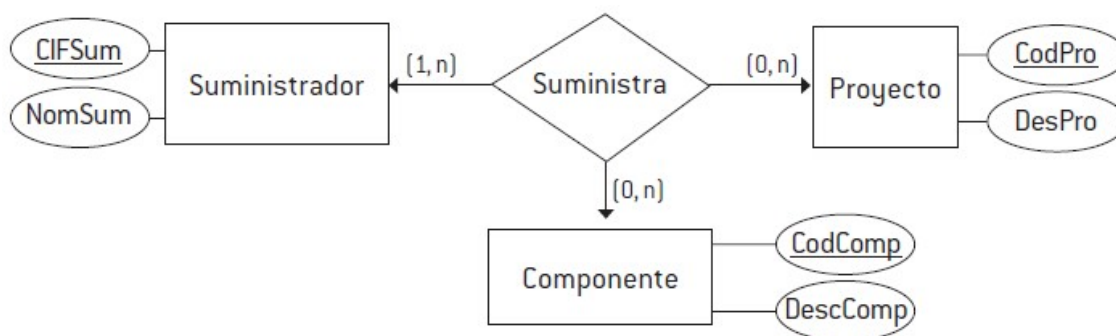


Figura 16: Representación de cardinalidades en una relación ternaria.

2.3.2. Atributos en relaciones.

Los atributos pueden pertenecer a las entidades, pero también a las relaciones. Es necesario asignar atributos a relaciones N:M en algunos casos, concretamente en aquellos en los que para cada par de ocurrencias de las entidades relacionadas es necesario registrar cierta información.

En el ejemplo de la figura 13, por cada pedido realizado a la empresa y por cada artículo incluido en dicho pedido, será necesario registrar el número de unidades que se solicitan de ese artículo en ese pedido (atributo *Unidades*). Este atributo no podemos asignárselo a la entidad *Pedido* porque en un pedido se pueden solicitar varios artículos y por

cada uno de ellos se puede solicitar un número de unidades distinto, y un atributo para una ocurrencia de una entidad no puede tomar varios valores. Tampoco podríamos asignárselo a la entidad *Artículo* porque un artículo se puede solicitar en varios pedidos y en cada uno de estos pedidos se pueden solicitar diferentes cantidades de cada artículo. Por tanto, el atributo *Unidades* debe ser asignado a la relación *incluye*, porque nos indica el número de unidades que se solicita por cada pareja de pedido-artículo, es decir, el número de unidades que de cada artículo se solicita en cada pedido.

En el esquema Entidad-Relación que aparece en la figura 15 también sería necesario añadir un atributo a la relación N:M presente si quisiésemos saber en cuántos componentes subordinados se descompone cada componente. Este atributo *Cantidad* nos indica por cada par componente superior – componente subordinado, de cuántas unidades de componente subordinado consta cada componente superior (figura 17).

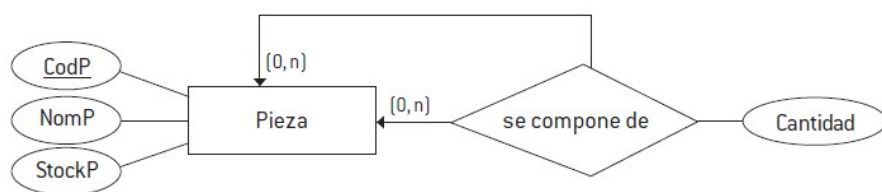


Figura 17: Relación N:M reflexiva con atributo.

También es posible asignar atributos a relaciones 1:1 o 1:N, pero en estos casos siempre es posible asignar el atributo también a alguna de las dos entidades participantes en la relación. Así, en el siguiente ejemplo el atributo *TiDir* (que indica el tipo de director de un departamento) se puede colocar en la relación *dirige*, pero también en la entidad *Departamento* porque todo departamento siempre tiene un director, del que es necesario conocer su tipo.

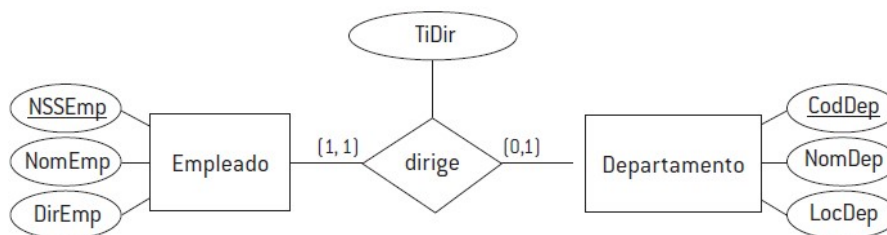


Figura 18: Relación 1:1 con atributo.

2.3.3. Entidades fuertes y débiles.

Aunque hasta ahora solo hemos hablado de entidades como tal, puede haber dos tipos de ellas:

- Entidades regulares o fuertes son aquellas para las cuales las ocurrencias de la entidad tienen existencia propia. Una entidad regular se representa como se he hecho hasta ahora, es decir, mediante un rectángulo en el interior del cual se coloca el nombre de la entidad en cuestión.
- Entidades débiles son aquellas para las cuales la existencia de una ocurrencia de la entidad débil depende de la existencia de una ocurrencia de la entidad regular de la que depende, de manera que si desaparece una ocurrencia de la entidad regular, desaparecerán automáticamente todas las ocurrencias de la entidad débil dependiente. Por ejemplo, la entidad *Libro* es una entidad regular, mientras que la entidad *Ejemplar* (ejemplar de libro) es una entidad débil que depende de la entidad *Libro* porque la existencia de un ejemplar de un libro depende de que exista el libro correspondiente. Una entidad débil se representa mediante dos rectángulos concéntricos en el interior de los cuales se coloca el nombre de la entidad en cuestión. Por ejemplo, la entidad *Ejemplar* se representa así:



Figura 19: Representación gráfica de una entidad débil.

Por otro lado, las relaciones pueden ser de dos clases en función del tipo de entidades que asocian:

- Relaciones regulares: Son aquellas que asocian entidades regulares.
- Relaciones débiles: Son aquellas que asocian una entidad débil con la entidad regular de la que depende.

Las relaciones débiles se representan frecuentemente por medio de dos rombos concéntricos, y la cardinalidad al lado de la entidad regular es casi siempre (1,1). Pues bien, estas relaciones débiles pueden ser de dos tipos:

- Dependencia en identificación: Se da este tipo de dependencia cuando la identificación de las ocurrencias de la entidad débil no se puede llevar a cabo con sus propios atributos, sino que se requiere para ello del AIP de la entidad regular correspondiente. Por ejemplo, si suponemos que los números de ejemplar de cada libro son números correlativos comenzando por el 1, el número de ejemplar no es único a nivel de toda la base de datos porque para todos los libros existirá el ejemplar número 1. Sin embargo, será única la pareja de atributos ISBN y número de ejemplar. Por ello, para identificar cada ejemplar de

un libro, se requiere el AIP del libro (ISBN) más un número de ejemplar (AIP de la propia entidad débil). Una dependencia en identificación se simboliza añadiendo las letras ID en el interior del doble rombo que representa la relación.

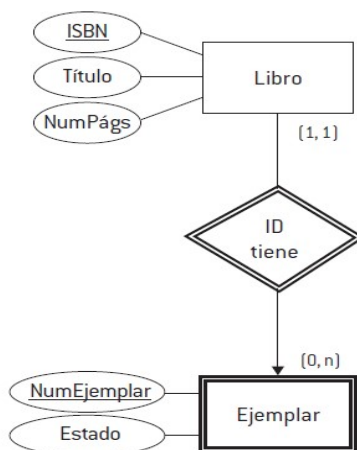


Figura 20: Dependencia en identificación.

- Dependencia en existencia: Cuando una relación débil no es una dependencia en identificación, se trata de una dependencia en existencia, algo que es intrínseco a todas las relaciones débiles, puesto que las ocurrencias de la entidad débil solo pueden existir si existe la ocurrencia de la entidad regular de la que dependen. Por ejemplo, la entidad *Provincia* depende en existencia de la entidad *Comunidad* (comunidad autónoma), dado que los datos de una provincia solo tienen sentido si en la base de datos se almacenan los datos de la comunidad autónoma a la que pertenece la provincia, y los datos de la provincia solo se almacenarán en la base de datos mientras estén los de la comunidad autónoma correspondiente. Una dependencia en existencia se representa incluyendo la letra E en el interior del doble rombo que representa la relación.

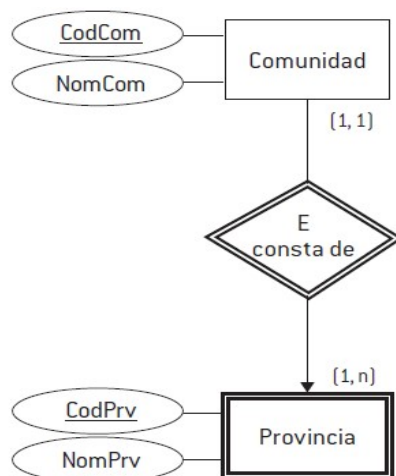


Figura 21: Dependencia en existencia.

2.3.4. Jerarquías de tipos y subtipos.

En muchas ocasiones, al analizar un universo del discurso, se detectan entidades similares que difieren en pocos atributos. En estos casos, puede ser recomendable establecer una jerarquía de tipos y subtipos por la cual se generalicen las entidades con atributos comunes (subtipos) en una entidad supertipo, la cual poseerá dichos atributos comunes. De esta forma, la entidad supertipo poseerá los atributos comunes a los subtipos, quedando para los subtipos solo los atributos específicos de cada uno de ellos. De manera similar, en el caso de las relaciones, aquellas que afectan a todos los subtipos se asocian al supertipo, dejando para los subtipos las específicas de cada uno de ellos.

Por ejemplo, en la figura 22, se ha creado un supertipo para los atributos comunes a las entidades *Profesor* y *Estudiante* (*NIF*, *Nombre*, *Teléfono*, *Dirección*), dejando en cada subtipo los atributos propios.

Este tipo de relaciones se representa mediante un triángulo invertido que generalmente lleva la leyenda “es un”, con la base situada hacia el supertipo y unido al supertipo y a los subtipos, siendo las cardinalidades siempre (1,1) en el supertipo y (0,1) o (1,1) en los subtipos.

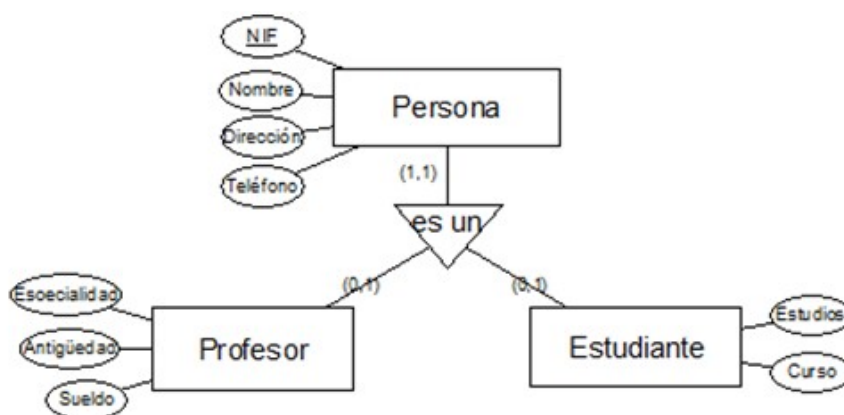


Figura 22: Jerarquía de tipos y subtipos.

Una característica muy importante de estas jerarquías es la herencia o propiedad por la cual los atributos del supertipo lo son también de los subtipos (son heredados por los subtipos). Así, tanto un profesor como un estudiante tienen NIF, nombre, dirección y teléfono, que son atributos heredados de su supertipo *Persona*. Además, los profesores tienen una especialidad, una antigüedad y un sueldo, mientras que los estudiantes cursan unos estudios y están realizando un determinado curso.

Las generalizaciones pueden ser de diferentes tipos:

- Una generalización será total si no hay ocurrencias en el supertipo que no pertenezcan a ninguno de los subtipos, lo que significaría en el caso del ejemplo propuesto, que todas las personas tienen que ser o profesores o estudiantes o ambas cosas a la vez. La generalización será parcial si hay ocurrencias en el supertipo que no pertenecen a ninguno de los subtipos. En nuestro caso, la jerarquía sería parcial si pudiese haber personas que no son ni profesores ni estudiantes. Una jerarquía total se representa dibujando un círculo en la línea que une el triángulo con el supertipo.
- Una generalización es exclusiva si un supertipo solo puede pertenecer a un subtipo o a ninguno de los subtipos presentes. En este caso, esto quiere decir que una persona solo puede ser o profesor o estudiante o ninguno de las dos cosas. Una generalización es solapada o superpuesta si un supertipo puede pertenecer a varios subtipos a la vez. Una generalización exclusiva se representa dibujando un arco que abarca las líneas que unen al triángulo con los subtipos.

Por lo que se ha indicado anteriormente, si consideramos que una persona tiene que pertenecer obligatoriamente a uno de los subtipos (generalización total), deberemos señalar este hecho dibujando un círculo. Si además consideramos que una persona solo puede ser o profesor o estudiante, pero no ambas cosas a la vez (generalización exclusiva), deberemos dibujar un arco, por lo que el esquema E-R nos quedaría como se ve en la figura 23.

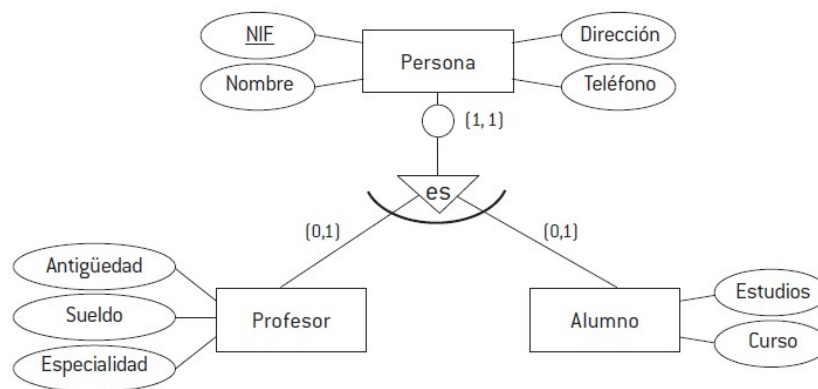


Figura 23: Jerarquía de tipos y subtipos total y exclusiva.

2.3.5. Control de la redundancia.

En ocasiones en los esquemas E-R aparecen relaciones redundantes que es aconsejable eliminar.

En un esquema E-R puede haber una relación redundante si hay un ciclo, es decir, el que exista un ciclo es una condición necesaria para la existencia de una relación redundante, pero esta condición no es suficiente, lo que quiere decir que a veces aunque exista un ciclo en un esquema E-R, puede no haber redundancias. Debe tenerse en cuenta lo siguiente:

- Las relaciones con atributos no se pueden eliminar, es decir, nunca son redundantes.
- Las relaciones débiles (dependencias en existencia o en identificación) tampoco se pueden eliminar.
- Para que cualquier otra relación se pueda eliminar, su eliminación no debe suponer una pérdida de semántica, lo que quiere decir, que para que una relación sea redundante, la información que nos proporciona la misma se debe poder obtener por medio de las relaciones que no se eliminan.

Consideremos el siguiente esquema E/R:

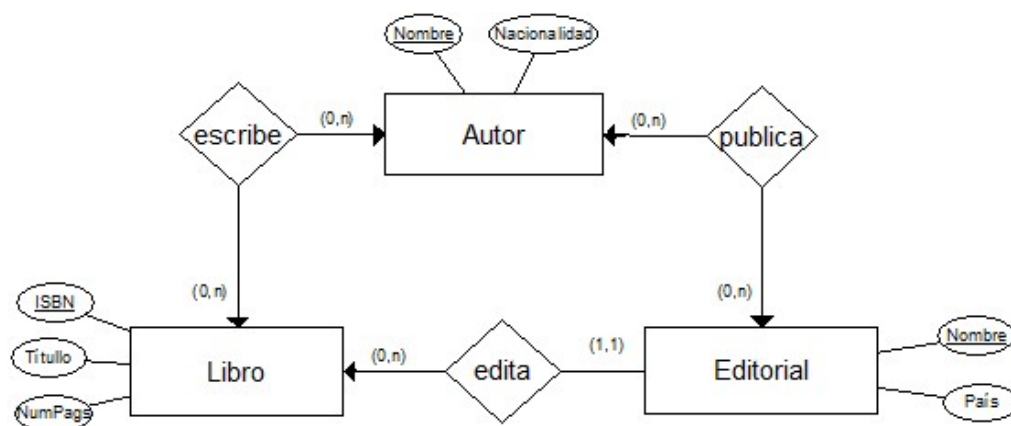


Figura 24: Relación redundante

Las relaciones *escribe* y *edita* no son redundantes porque la información que nos proporcionan no puede ser conseguida de ninguna manera por medio de las relaciones que quedarían al eliminarlas. Con respecto a la relación *publica*, será redundante si la información que nos proporciona se puede obtener aun eliminándola por medio de las otras relaciones. Veamos si esto es posible.

- Aunque eliminemos la relación *publica*, es necesario seguir sabiendo para qué editoriales ha publicado libros un autor. Pues bien, si a partir de un autor conozco los libros escritos por este y por cada libro la editorial que lo ha publicado, sé también en qué editoriales ha publicado libros dicho autor.
- En el sentido contrario, debe ser posible a partir de una editorial conocer los autores que han publicado libros para ella. Pues bien, si a partir de una editorial sé qué libros ha

editado y para cada uno de estos libros puedo saber sus autores, también puedo conocer todos los autores que han publicado libros para dicha editorial.

En consecuencia, la relación *publica* es redundante y por lo tanto sería aconsejable eliminarla del esquema E-R.