

U.T.2. DISEÑO CONCEPTUAL. MODELO ENTIDAD/RELACIÓN

1. REPRESENTACIÓN DEL PROBLEMA

Una base de datos representa la información contenida en alguna organización del mundo real. El **diseño de bases de datos** consiste en extraer todos los datos relevantes de un problema, por ejemplo, saber qué datos están implicados en el proceso de facturación de una empresa o qué datos son necesarios para llevar el control médico de los pacientes de un hospital.

Para extraer estos datos se debe realizar un análisis en profundidad del dominio del problema y determinar qué datos son útiles para la base de datos, descartando los que no lo son. Una vez extraídos estos datos esenciales comienza el proceso de *modelización*, esto es, mediante una herramienta de diseño de bases de datos, construir un *esquema* que represente toda esa información que se requiere almacenar.

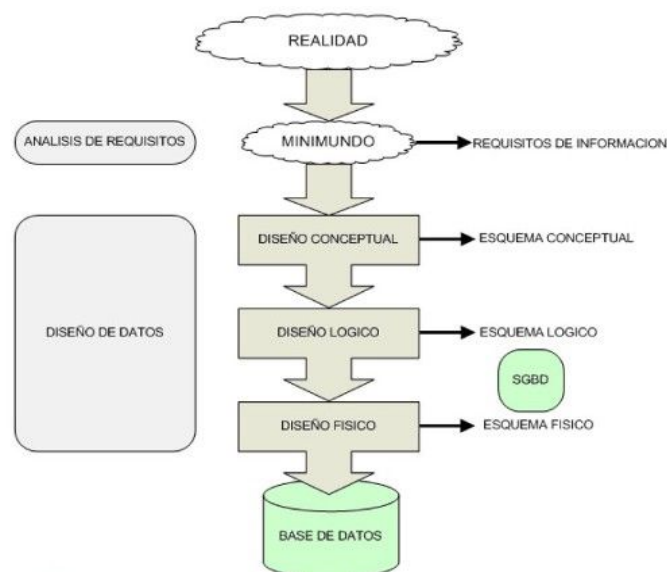
Un *esquema* es una representación de una parte específica de la realidad, creada usando un determinado modelo de datos.

2. DISEÑO DE BASES DE DATOS

El diseño de Base de Datos desempeña un papel central en el empleo de los recursos de información en la mayoría de las organizaciones.

Se han desarrollado metodologías y técnicas de diseño y se ha alcanzando un consenso sobre la descomposición del proceso en fases, sobre los principales objetivos de cada fase y sobre las técnicas para conseguir estos objetivos.

Así, el diseño de una Base de Datos se descompone en : Diseño Conceptual, Diseño Lógico y Diseño Físico:



- El **diseño conceptual** parte de las especificaciones de requisitos de usuario y su resultado es el esquema conceptual de la Base de Datos. Un esquema conceptual es una descripción de alto nivel de la estructura de la Base de Datos, independientemente del SGBD que se vaya a utilizar. Los procesos de definición de requisitos y del diseño conceptual exigen identificar las exigencias de información de los usuarios y representarlos en un modelo bien definido.

3. EL MODELO ENTIDAD/RELACIÓN

En este punto trataremos el diseño conceptual de una base de datos mediante el modelo ER. Lo que explicaremos es aplicable al diseño de cualquier tipo de bases de datos -relacional, jerárquica, etc.-, porque, como ya hemos dicho, en la etapa del diseño conceptual todavía no se tiene en cuenta la tecnología concreta que se utilizará para implementar la base de datos.

El modelo ER es uno de los enfoques de modelización de datos que más se utiliza actualmente por su simplicidad y legibilidad. Su legibilidad se ve favorecida porque proporciona una notación diagramática muy comprensiva. Es una herramienta útil tanto para ayudar al diseñador a reflejar en un modelo conceptual los requisitos del mundo real de interés como para comunicarse con el usuario final sobre el modelo conceptual obtenido y, de este modo, poder verificar si satisface sus requisitos.

El nombre completo del modelo ER es *entity-relationship*, y proviene del hecho de que los principales elementos que incluye son las entidades y las interrelaciones (entities y relationships). Traduciremos este nombre por ‘entidad-interrelación’.

El origen del modelo ER se encuentra en trabajos efectuados por Peter Chen en 1976. Posteriormente, muchos otros autores han descrito variantes y/o extensiones de este modelo. Así pues, en la literatura se encuentran muchas formas diferentes del modelo ER que pueden variar simplemente en la notación diagramática o en algunos de los conceptos en que se basan para modelizar los datos.

4. ELEMENTOS DEL MODELO E/R

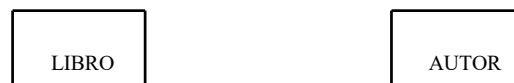
En el modelo E/R se puede distinguir como elementos fundamentales las *entidades*, los *atributos* y las *relaciones*.

4.1. Entidad

Se puede definir la entidad como cualquier objeto (real o abstracto) acerca del cual queremos almacenar información en la base de datos.

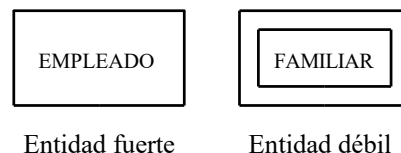
Las entidades pueden ser de distinta naturaleza: objetos físicos (coches, libros, ...), personas (clientes, empleados, médicos, ...), lugares (ciudades, provincias, ...), organizaciones (hospitales, empresas), etc.

La representación gráfica de este objeto es un rectángulo etiquetado con el nombre del tipo de entidad, como podemos ver en la figura.



Existen dos clases de entidades: *regulares* (o fuertes), que son aquellas que tienen existencia por sí mismas, como LIBRO y AUTOR, y *débiles*, cuya existencia depende de otro tipo de entidad, por ejemplo, FAMILIAR depende de EMPLEADO, y la desaparición de un empleado de la BD hace que desaparezcan también todos los familiares que estaban a su cargo.

Los tipos de entidad débil se representan con dos rectángulos concéntricos con su nombre en el interior, como se ve en la figura.



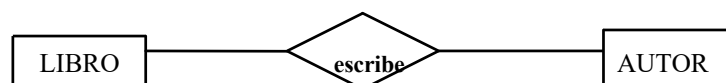
Ocurrencia de una entidad

Es una instancia de una determinada entidad, es decir, una unidad del conjunto que representa la entidad. La entidad empleado tiene varias instancias que son cada uno de los empleados de la empresa, como por ejemplo, Jesús Corbalán.

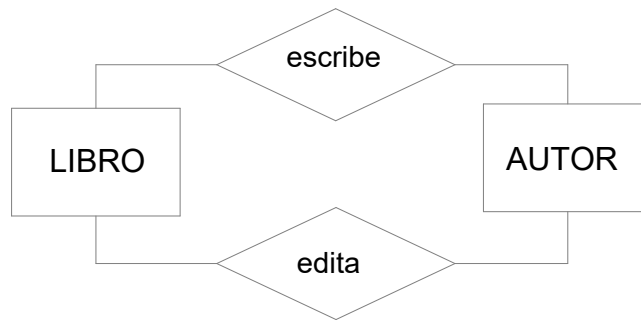
4.2. Relación

Se entiende por relación o interrelación aquella asociación o correspondencia existente entre entidades.

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. Generalmente los nombres de las relaciones corresponden a verbos, pues las relaciones suelen describir acciones entre dos o más entidades.



Entre dos tipos de entidad puede existir más de un tipo de interrelación.

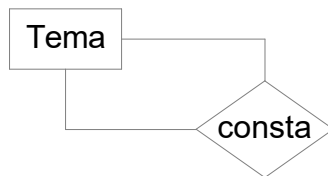


Elementos de un tipo de interrelación

Nombre: como todo objeto del modelo E/R, cada tipo de relación tiene un nombre que lo distingue unívocamente del resto y mediante el cual ha de ser referenciado.

Grado: es el número de tipos de entidad que participan en un tipo de interrelación.

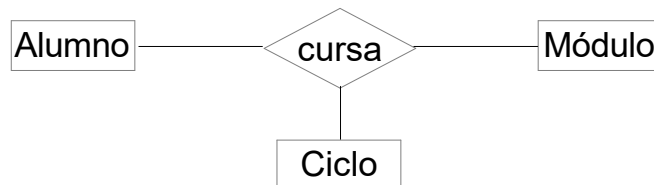
Grado 1 (relaciones unarias o reflexivas): la misma entidad participa más de una vez en la relación.



Grado 2 (relaciones binarias): son dos las entidades que participan en la relación. Son las más comunes.



Grado 3 (relaciones ternarias): en la relación participan tres entidades.



Grado n (relaciones n-arias donde $n > 3$): en la relación participan más de 3 entidades. Aparecen muy raras veces ya que generalmente se pueden descomponer en varias relaciones de grado 2 o de grado 3.

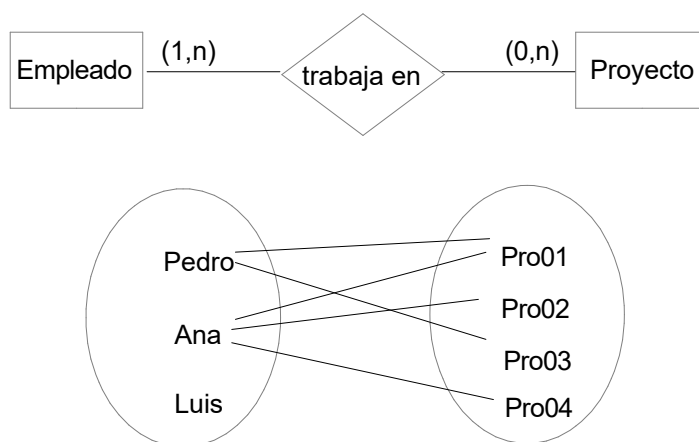
Nota: Si en tu diagrama entidad relación aparecen relaciones de grado > 3 , es posible que la interpretación de la realidad sea incorrecta. Incluso si aparecen relaciones de grado 3 **intenta** descomponerlas en varias de grado 2, siempre que no se pierda semántica.

Participación: indica mediante una pareja de números el número máximo y mínimo de ocurrencias de un tipo de entidad que pueden estar interrelacionadas con una ocurrencia

del otro tipo, u otros tipos de entidad que participan en el tipo de interrelación y que aparecen. Su representación gráfica es una etiqueta del tipo (0,1), (1,1), (0,n) o (1,n), según corresponda.

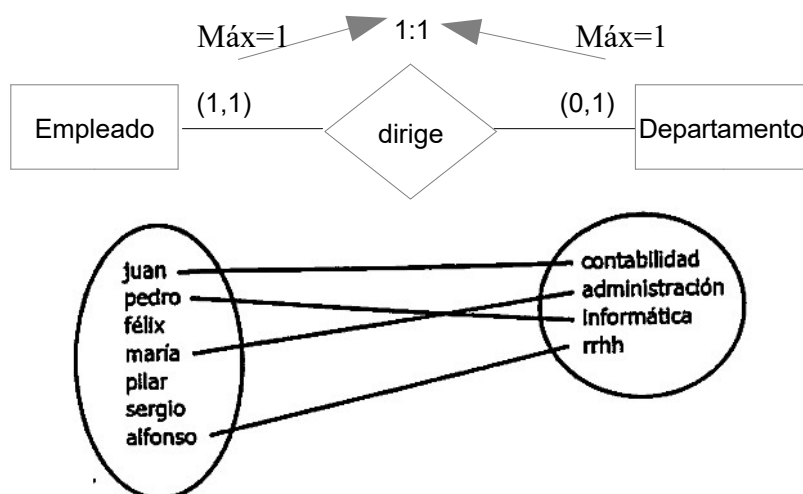
Participación	Significado
(0,1)	Mínimo cero, máximo uno
(1,1)	Mínimo uno, máximo uno
(0,n)	Mínimo cero, máximo n (muchos)
(1,n)	Mínimo uno, máximo n (muchos)

Las reglas que definen la participación de una ocurrencia en una relación son las *reglas de empresa*, es decir, se reconocen a través de los requisitos del problema. Por ejemplo, los empleados pueden trabajar en varios proyectos o en ninguno (cuando están de vacaciones). En un proyecto puede trabajar mínimo 1 empleado, aunque puede haber más. Esto se representaría así:



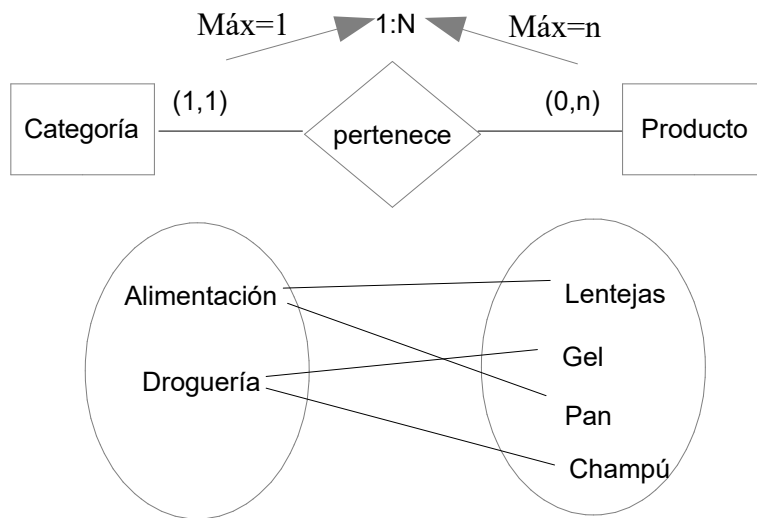
Cardinalidad: es el número máximo de ocurrencias de cada tipo de entidad que pueden intervenir en una ocurrencia del tipo de interrelación que se está tratando; para representarlo gráficamente se pone etiqueta como 1:1, 1:N o N:M, según corresponda.

Cardinalidad 1:1



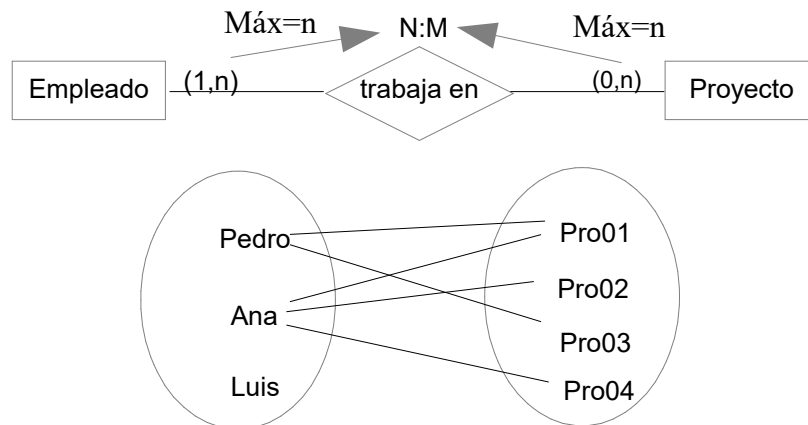
La cardinalidad 1:1 de esta relación establece que un empleado o no es jefe o lo es de un solo departamento y un departamento sólo puede tener un empleado jefe.

Cardinalidad 1:N



La cardinalidad 1:N de esta relación especifica que un producto solo puede pertenecer a una categoría y que a una categoría pueden pertenecer varios productos.

Cardinalidad M:N

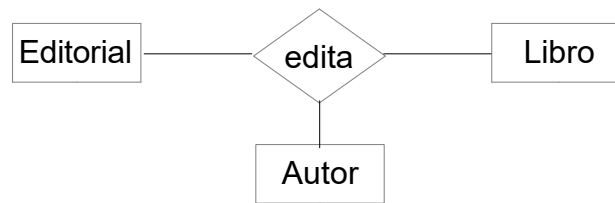


La cardinalidad M:N (muchos a muchos) de esta relación permite que un empleado pueda trabajar en varios proyectos y que en un mismo proyecto puedan trabajar varios empleados.

Cardinalidad en una relación ternaria

Para calcular la cardinalidad de una relación ternaria se tomará una de las tres entidades y se combina con las otras dos. A continuación, se calcula la participación de la entidad en la combinación de las otras dos. Posteriormente, se hará lo mismo con las otras dos entidades. Finalmente, tomando los máximos de las participaciones se generan las cardinalidades.

Lo veremos mediante el siguiente ejemplo:



Hay que responder a las siguientes preguntas:

¿Cuántos autores puede tener un determinado libro publicado en una determinada editorial?

Mínimo 1, máximo n. Participación de Autor (1,n).

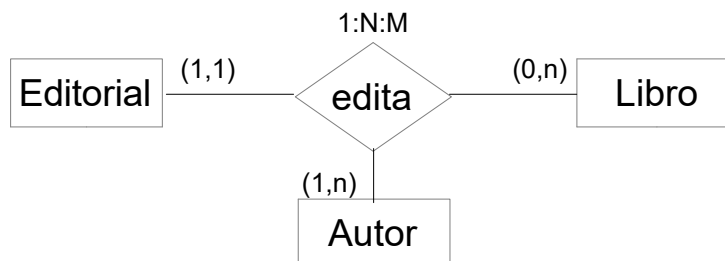
¿Cuántos libros puede tener un determinado autor publicados en una determinada editorial?

Mínimo 0, máximo n. Participación de Libro (0,n).

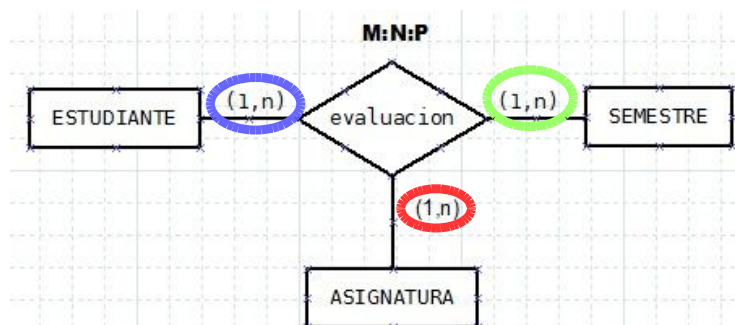
¿En cuántas editoriales puede un determinado autor publicar un determinado libro?

Mínimo 1, máximo 1. Participación de Editorial (1,1).

Tomando los máximos de cada participación se obtiene que la cardinalidad de la relación es 1:N:M.



Veamos otro ejemplo.



Un estudiante en un semestre se puede evaluar de 1 o varias asignaturas (1,n).

Un estudiante de una asignatura se puede evaluar en 1 o varios semestres (1,n).

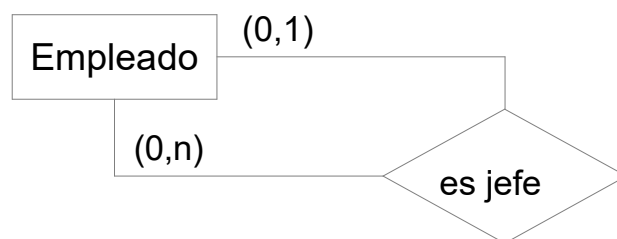
En un semestre y de una asignatura se pueden evaluar 1 o varios estudiantes (1,n).

Cardinalidad de las relaciones reflexivas

En las relaciones reflexivas la misma entidad juega dos papeles distintos en la relación. Para calcular su cardinalidad hay que extraer las participaciones según los dos roles existentes. Por ejemplo, en la relación reflexiva “Es jefe”, la entidad Empleado aparece con dos roles. El primer rol es el empleado como jefe y el segundo rol es el empleado como subordinado. Así se puede calcular las participaciones preguntando:

- ¿Cuántos subordinados puede tener un jefe? Un jefe puede tener un mínimo de 0 (empleados que no son jefes) y un máximo de n: (0,n)
- ¿Cuántos jefes puede tener un subordinado? Un mínimo de 0 (un empleado sin jefes sería el responsable de la empresa) y un máximo de 1 (suponiendo una estructura típicamente jerárquica): (0,1)

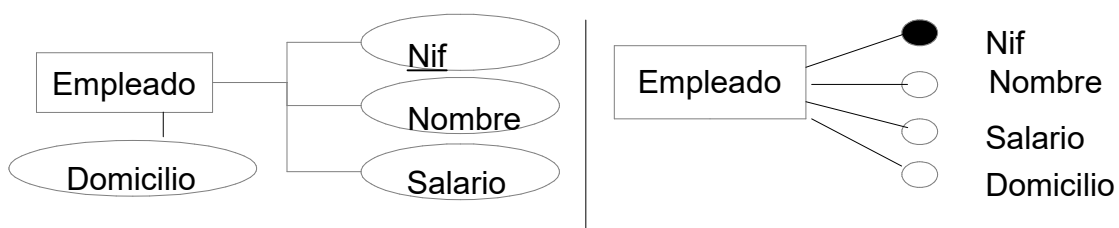
Por lo tanto, la relación tendría la cardinalidad 1:N



5.3. Atributos y dominios

Cada una de las propiedades o características que tiene un tipo de entidad o un tipo de interrelación se denomina ATRIBUTO.

Se representación gráficamente mediante elipses conectadas directamente a la entidad, como en la figura de la izquierda. Cuando la entidad tiene muchos atributos y/o hay muchas entidades y relaciones en el diagrama Entidad Relación, por razones de espacio, se pueden representar como en la figura de la derecha.



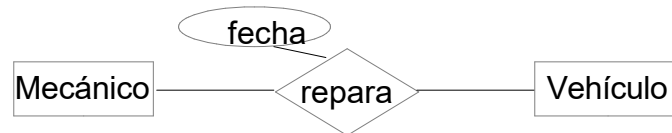
Atributo clave

Entre todos los atributos de un tipo de entidad debemos elegir uno o varios que identifiquen unívocamente cada una de las ocurrencias de ese tipo de entidad. Este atributo o conjunto de atributos se denomina **atributo clave**. En el ejemplo anterior, Nif sería el atributo clave de la entidad Empleado, pues no hay dos personas con el mismo Nif. Por tanto, si especificamos un Nif concreto estaremos refiriéndonos a un único empleado.

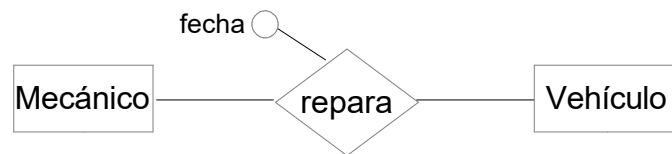
Se suele representar mediante subrayado (figura de la izquierda) o rellenando el círculo correspondiente a dicho atributo (figura de la derecha).

Atributos de relación

Los tipos de interrelación (relaciones) también pueden tener atributos. El atributo fecha de la relación repara del ejemplo siguiente corresponde a la reparación de un determinado vehículo por un determinado mecánico. La fecha de reparación no podría asignarse a la entidad mecánico, porque un mecánico realizará muchas reparaciones en diferentes fechas. Tampoco podríamos asignar la fecha de reparación a la entidad vehículo, ya que un mismo vehículo podrá ser reparado en diferentes fechas (diferentes reparaciones):



O:



Dominios

Cada una de las características que tiene una entidad pertenece a un dominio. El dominio representa la naturaleza del dato, es decir, si es un número entero, una cadena de caracteres o un número real. Incluso naturalezas más complejas, como una fecha o una hora. Por ejemplo, los siguientes atributos de la entidad empleado pertenecen a los siguientes dominios:

Atributo	Dominio
Nif	Cadena de caracteres de longitud 10
Nombre	Cadena de caracteres de longitud 50
Salario	Números reales
Domicilio	Cadena de caracteres de longitud 50
Fecha_nacimiento	Fecha

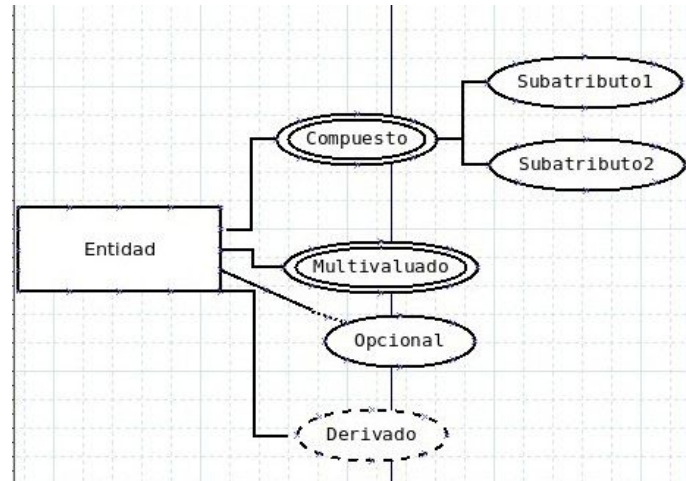
Tipos de atributos

Se pueden clasificar los atributos según las siguientes restricciones:

- **Atributos obligatorios:** Un atributo debe tomar un valor obligatoriamente.
- **Atributos opcionales:** Un atributo puede no tomar un valor porque sea desconocido. En este caso, el atributo tiene el valor nulo.
- **Atributos compuestos:** Un atributo compuesto es aquel que se puede descomponer en atributos más sencillos, por ejemplo, el atributo hora_de_salida se puede descomponer en dos (hora y minutos).
- **Atributos univaluados:** Un atributo que toma un único valor.
- **Atributos multivaluados:** Estos atributos pueden tomar varios valores, por ejemplo el atributo teléfono puede tomar los valores de un teléfono móvil y un teléfono fijo.

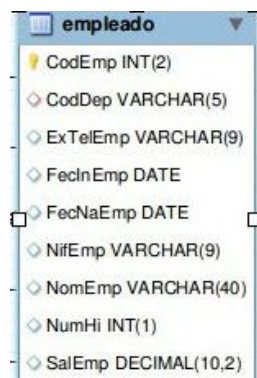
- **Atributo derivado:** Son aquellos cuyo valor se puede calcular a través de otros atributos. Por ejemplo, el atributo Edad se puede calcular a partir de la fecha de nacimiento de una persona.

La notación más extendida entre los diseñadores de bases de datos para representar los tipos de atributos es la siguiente:



Otras notaciones para entidades y atributos

Aunque existen multitud de notaciones para representar entidades y atributos, es la notación original de Peter Chen la más usada. Sin embargo, por simplificar la construcción de esquemas a través de herramientas software, se opta por usar otras notaciones. Por ejemplo, la herramienta MySQL Workbench utiliza una sintaxis muy similar a la que usa la notación UML para representar las características de un objeto:



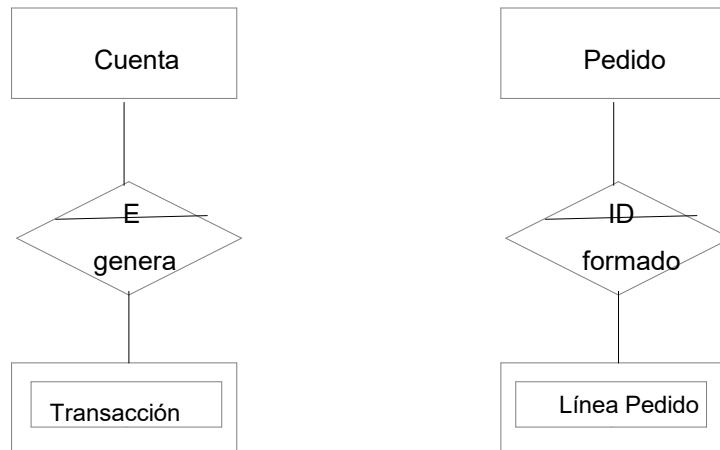
6. DEPENDENCIA EN EXISTENCIA Y EN IDENTIFICACIÓN

Como en el caso de los tipos de entidad, los tipos de interrelación se clasifican también en **regulares** y **débiles**, según estén asociando dos tipos de entidad regulares, o un tipo de entidad débil con un tipo de entidad regular, respectivamente.

Hay que distinguir dentro del tipo de interrelación débil la dependencia en existencia y la dependencia en identificación. Se dice que hay **dependencia en existencia** cuando las ocurrencias de un tipo de entidad (entidad débil) no pueden existir si desaparece la ocurrencia de la entidad regular de la cual dependen. La **dependencia es en identificación** cuando además de cumplirse la condición anterior, las ocurrencias del

tipo de entidad débil no se pueden identificar unívocamente mediante los atributos propios de la misma y exigen añadir la clave del tipo de entidad regular del cual dependen. Una dependencia en identificación es siempre una dependencia en existencia (no ocurre lo contrario).

Si la dependencia es en identificación, el rombo que representa la interrelación va etiquetado con **ID**, y con una **E** en caso de que la dependencia sea sólo en existencia.



Las transacciones que se dan en una cuenta bancaria no tienen sentido si no existe la cuenta a la que están asociadas. Éste sería un ejemplo de entidad débil con dependencia en existencia.

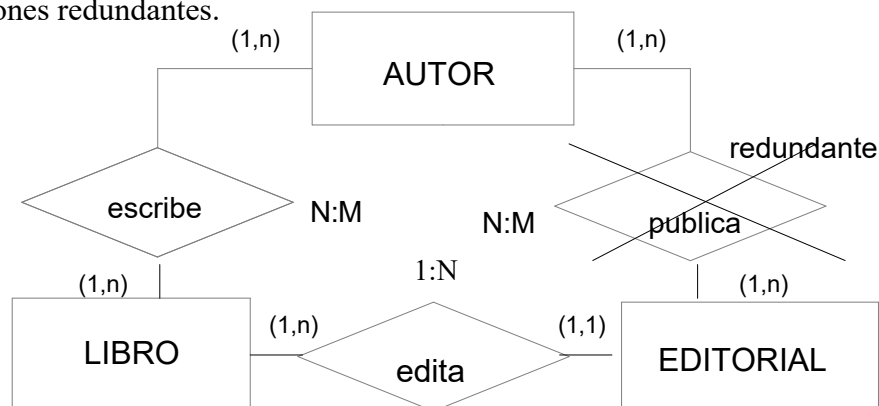
Las líneas de pedido, además de depender en existencia del pedido a que corresponden necesitan el atributo clave del pedido para poder identificarse. Éste sería un ejemplo de entidad débil con dependencia en identificación.

Nosotros sólo representaremos las entidades débiles en identificación (ID). Las entidades débiles en existencia, como Transacción, las representaremos como entidades fuertes o regulares (sin doble rectángulo).

7. CONTROL DE REDUNDANCIA EN LOS ESQUEMAS E/R

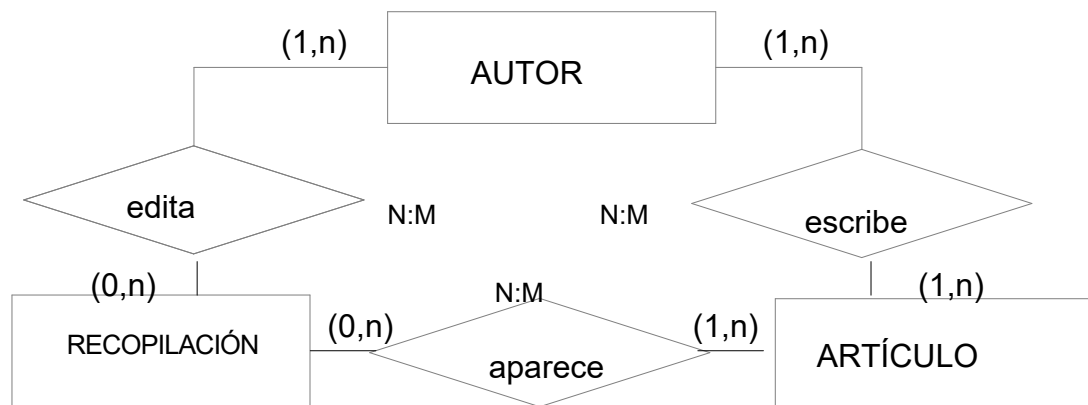
Una vez construido un esquema E/R hay que analizar si se presentan redundancias.

Además de la existencia de atributos redundantes, como los que se derivan de otros mediante algún cálculo y que deben ser eliminados del esquema E/R, hay que estudiar detenidamente los ciclos en el diagrama E/R, ya que *pueden* indicar la existencia de interrelaciones redundantes.



En el ejemplo anterior se da un ciclo entre AUTOR, LIBRO y EDITORIAL, por lo que en principio es posible que aparezca alguna interrelación redundante. En efecto, si se conocen los libros de un autor y las editoriales que los han editado, se puede deducir fácilmente en qué editoriales ha publicado dicho autor; de forma análoga, dada una editorial, si sabemos qué libros ha publicado, podemos deducir qué autores han escrito para ella, por lo que la interrelación **publica** entre las entidades AUTOR y EDITORIAL es redundante.

Sin embargo, en esta otra figura, a pesar de existir un ciclo, no hay ninguna interrelación redundante. En lo que respecta a la interrelación **escribe**, no puede deducirse de las otras dos, ya que aunque sepamos las recopilaciones que ha editado un autor y los artículos que han aparecido en ellas, no podemos saber los artículos que ha escrito dicho autor. La interrelación **edita** tampoco es redundante, ya que un autor escribe varios artículos que pueden aparecer en recopilaciones sin que eso implique que dicho autor edite esas recopilaciones. Por último, la interrelación **aparece** tampoco es redundante, ya que un artículo escrito por un autor no tiene porqué aparecer necesariamente en las recopilaciones que éste edite: puede aparecer en otras recopilaciones o incluso en revistas, obsérvese la cardinalidad mínima cero en la entidad RECOPIACIÓN con respecto a la entidad ARTÍCULO en la interrelación **aparece**.



Existen además otros casos en los que la interrelación, a pesar de poder ser deducida a partir de otras presentes en el esquema, no se puede eliminar porque posee atributos.

Se puede decir, como norma general, que la existencia de un ciclo no implica la existencia de interrelaciones redundantes

8. CONSTRUCCIÓN DE UN DIAGRAMA E/R

A continuación se presenta una guía metodológica para crear un diagrama entidad relación a partir de un análisis de requisitos:

1. Leer varias veces el problema.
2. Obtener una lista inicial de candidatos a entidades, relaciones y atributos. Se realiza siguiendo los siguientes consejos:
 - Identificar las entidades. Suelen ser aquellos nombres comunes que son importantes para el desarrollo del problema. Por ejemplo, empleado, vehículo, agencia, etc.

- No hay que obsesionarse en los primeros pasos por distinguir las entidades fuertes de las débiles.
 - Extraer los atributos de cada entidad, identificando aquellos que pueden ser clave. Se suelen distinguir por ser adjetivos asociados a un nombre común seleccionado como entidad. Por ejemplo, color (de un vehículo).
 - Identificar los atributos de las relaciones, si es que los hay. Suelen ser adjetivos también pero aplicables a la relación y no a ninguna de las entidades relacionadas.
 - Si un nombre común aparece pero no tenemos información de él, podrá ser un atributo de otra entidad. Por ejemplo, si aparece el autor de un libro, pero no tenemos información adicional (fecha de nacimiento, nacionalidad, ...) será un atributo de la entidad libro.
 - Identificar las relaciones. Se pueden ver extrayendo los verbos del texto del problema. Por ejemplo: agente inmobiliario vende edificio. En este caso, agente inmobiliario y edificio serían las entidades y *vende* sería la relación.
3. Averiguar las participaciones y cardinalidades. Generalmente se extraen del propio enunciado del problema. Si no vienen especificadas, se elegirá la que almacene mayor cantidad de información en la base de datos.
 4. Poner todos los elementos listados en el paso 2 en un esquema (diagrama) y volver a considerar la pertenencia de cada uno de los elementos listados a su categoría.
 5. Refinar el diagrama hasta que se eliminen todas las incoherencias posibles repasando los pasos anteriores en caso de encontrar algún atasco o concepto dudoso.