
TEMA 3. Diseño Conceptual de bases de datos relacionales.

1. El diseño conceptual

Simplificando lo comentado en el tema anterior y una vez especificados los requisitos de usuario que nuestra base de datos debe cumplir, el proceso de diseño de una base de datos debe distinguir, al menos, tres grandes fases:

- Diseño conceptual, cuyo objetivo es obtener una representación de la información con independencia de usuarios y aplicaciones en particular, y fuera de consideraciones sobre la eficiencia del ordenador.
- Diseño lógico, cuyo objetivo es transformar el diseño conceptual obtenido y adaptarlo al modelo de datos en el que se apoya el SGBD que se va a utilizar. En nuestro caso, el SGBD es relacional, por lo cual nos referiremos a este modelo de datos.
- Diseño físico, cuyo objetivo es conseguir una instrumentación lo más eficiente posible del diseño lógico.

Dado que el objetivo de este curso es el de introducir los conceptos básicos de las bases de datos, nos centraremos sólo en el diseño conceptual y el diseño lógico ya que el diseño físico depende de cada SGBD y de cada computadora en particular.

De forma práctica y expresado brevemente, el diseño conceptual consiste en extraer la información referente al trabajo de la empresa para la cual se desea desarrollar la base de datos, es decir, extraer aquellas entidades y acciones que son de uso habitual en la misma y que van a formar parte de la base de datos, y representarlo en forma de modelo. Para ello, la forma habitual de diseño es haber realizado previamente un análisis de los requisitos de usuario mediante la consulta con los empleados de la empresa, pues a partir de la misma se habrá de obtener el conjunto de entidades que van a formar parte de la base de datos, así como las acciones relevantes que pueden afectar al diseño de la base de datos. A partir de esta información podemos obtener el diseño conceptual donde se incluye la cardinalidad entre las entidades, que se obtiene en base a las posibilidades de relación existentes entre las entidades.

En este tema aprenderemos a realizar el diseño conceptual mediante el modelo entidad-relación (E/R), uno de los modelos de datos conceptuales más extendidos en las metodologías de diseño de bases de datos, obteniendo así un esquema para la base de datos. Con este esquema ya tenemos identificados los elementos y las relaciones entre ellos así como sus cardinalidades, y podremos pasar al diseño lógico.

2. El Modelo Entidad-Relación (E/R)

El modelo E/R fue propuesto por Peter P. Chen en 1976. Según Chen, *‘el modelo E/R puede ser usado como una base para una vista unificada de los datos’*, adoptando *‘el enfoque más natural del mundo real que consiste en entidades y relaciones’*.

Posteriormente, otros autores han investigado y escrito sobre el modelo, proponiendo importantes aportaciones, por lo que realmente no se puede considerar que exista un único modelo E/R, sino más bien lo que podríamos llamar una familia de modelos.

Como ya hemos comentado en sucesivas ocasiones, el modelo E/R está centrado en dos conceptos fundamentales: el de *entidad* y el de *relación* (los cuales detallaremos a continuación). Desde que fue propuesto, este modelo ha tenido una gran difusión y ha despertado un enorme interés en la comunidad informática dedicada a las bases de datos. En 1988 el ANSI lo seleccionó como modelo estándar para los sistemas de diccionarios de recursos de información.

2.1. Conceptos fundamentales

En el modelo E/R se pueden distinguir como elementos fundamentales las entidades, los atributos, y las relaciones, además del conjunto de valores.

Entidad

Se puede definir entidad como aquel objeto (real o abstracto) acerca del cual queremos almacenar información en la base de datos. Denominaremos a la estructura genérica en su sentido abstracto *tipo de entidad*, mientras que *entidad* será cada una de las ocurrencias o instancias de este tipo de entidad.

La representación gráfica de un tipo de entidad es un rectángulo etiquetado con el nombre del tipo de entidad:



Tres reglas generales que debe cumplir cualquier entidad son (Tardieu):

- Tiene que tener existencia propia.
- Cada ocurrencia de un tipo debe poder distinguirse de las demás.
- Todas las ocurrencias de un tipo de entidad deben tener los mismos tipos de características (atributos).

Existen dos clases de entidades: *regulares*, que tienen existencia por ellas mismas, y *débiles*, cuya existencia depende de otro tipo de entidad. Como ejemplo de entidad débil podríamos pensar en la entidad PAGO como los distintos pagos de un préstamo (el primer pago, el segundo,...) que dependerían de la entidad PRESTAMO, y la eliminación de PRESTAMO en nuestra base de datos obligaría a eliminar PAGO.

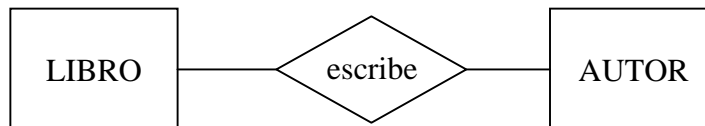
Los tipos de entidad débil se representan con dos rectángulos concéntricos con su nombre en el interior:



Relación

Se entiende por relación a aquella asociación o correspondencia existente entre entidades. Llamaremos *tipo de relación* a la estructura genérica del conjunto de relaciones existentes entre dos o más tipos de entidad.

El tipo de relación se representa mediante un rombo etiquetado con el nombre de la relación, unido mediante líneas a los tipos de entidad que asocia. Por ejemplo:

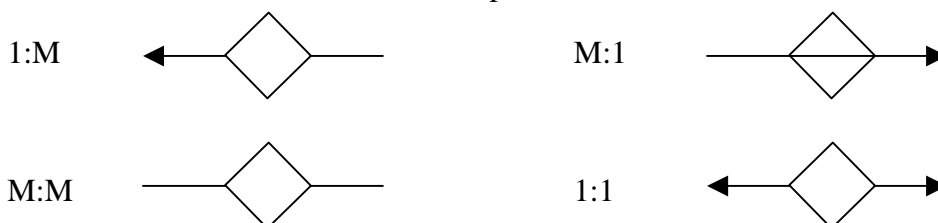


Entre dos tipos de entidades puede existir más de un tipo de relación.

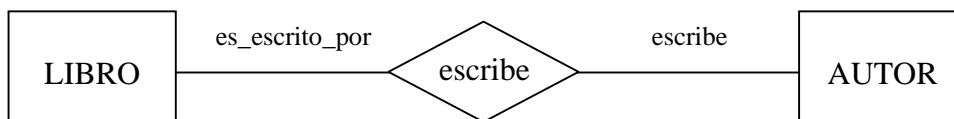
Una relación se define por su *nombre* y por su *grado*. El nombre es el identificador que se le da a la propia relación, y el grado equivale al número de tipos de entidad a los que asocia o relaciona. Así por ejemplo, en el caso anterior, ‘escribe’ es una relación de grado 2. También pueden existir relaciones de grado 1 (o reflexivas), de grado 3, 4, ...

En general, las relaciones de grado superior a 2 se suelen reducir a relaciones de grado 2, si la semántica de la relación lo permite, puesto que se facilita tanto la comprensión del diagrama como la posterior conversión a otros modelos.

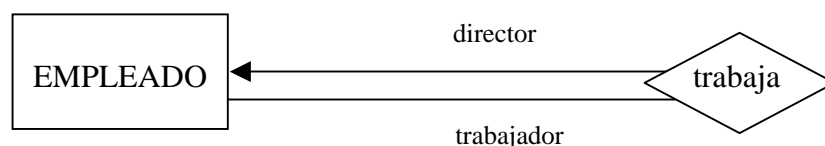
Otro elemento que caracteriza a las relaciones es el *tipo de correspondencia*, que es el número máximo de ocurrencias de cada tipo de entidad que pueden intervenir en una ocurrencia del tipo de relación que se está tratando. Gráficamente, esto se representa con alguna de estas etiquetas textuales: 1:1, 1:N, N:M. También se puede representar la correspondencia haciendo que las líneas que unen las entidades y la relación sean flechas, es decir, las distintas relaciones se representarían como:



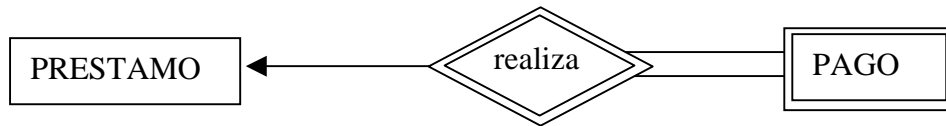
El *papel* o *rol* es la función que cada tipo de entidad realiza en el tipo de relación. Se representa con su nombre sobre la línea que une el tipo de entidad con el tipo de relación. Por ejemplo:



Normalmente no se suele escribir el *papel* que juega cada tipo de entidad en la relación, sólo para resaltar dicha información o evitar confusiones, como por ejemplo el de la entidad EMPLEADO y su JEFE inmediato (el cual a su vez es un empleado).



Por otra parte, si la línea que une una entidad con la relación asociada es doble en lugar de simple, implica que todas las entidades participan de al menos una relación. Por ejemplo, en el caso de PAGO y PRESTAMO, como PAGO es una entidad débil su existencia queda condicionada a la entidad PRESTAMO, por lo que necesariamente debe existir un préstamo para que pueda asociársele un pago así que lo representaríamos con doble línea. A su vez, a la relación que asocia ese pago con el préstamo correspondiente se le representa con doble rombo.



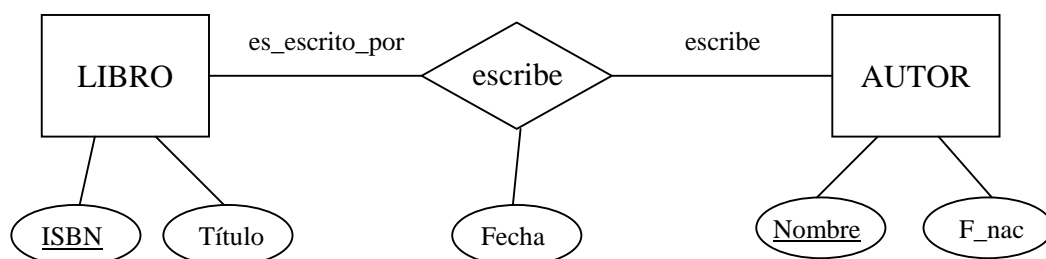
Atributo

Cada una de las propiedades o características que tiene un tipo de entidad o un tipo de relación se denomina ATRIBUTO. La representación gráfica de un atributo consiste en una elipse con el nombre del atributo en su interior.

De entre todos los atributos de un tipo de entidad debemos elegir uno o varios de forma que me identifiquen unívocamente a la entidad para que sean clave primaria. Por lo tanto, definimos los siguientes conceptos:

- **Superclave:** es el conjunto de uno o más atributos que juntos permiten identificar de forma única una entidad.
- **Clave candidata:** es una superclave para la cual no existe ningún subconjunto suyo que sea a su vez superclave.
- **Clave primaria:** es la clave candidata elegida por el diseñador de la base de datos para identificar a la entidad.

Los atributos que actúan como clave primaria se representarán de la misma forma que el resto de atributos, es decir mediante elipses, pero con el nombre del atributo subrayado. Veamos un ejemplo:



Para el caso de un tipo de entidad débil, y por definición de lo que es entidad débil, no tenemos clave primaria de entre sus atributos pero también necesitamos poder distinguir cada una de las entidades que dependen de la entidad fuerte. Al atributo o conjunto de atributos que nos permiten hacer esta distinción se le llama **discriminante** o **clave parcial**. La clave primaria del tipo entidad débil se forma con la clave primaria del tipo entidad fuerte más el discriminante del tipo entidad débil. Por otra parte, la forma de representar el discriminante es subrayando el nombre con una línea discontinua.

Por todo lo comentado hasta ahora, podemos deducir que en un tipo de entidad débil el tipo de relación que tiene sentido es 1:M y además, dicha relación no tendrá atributos descriptivos dado que estos podrían formar parte de los atributos del tipo entidad débil.

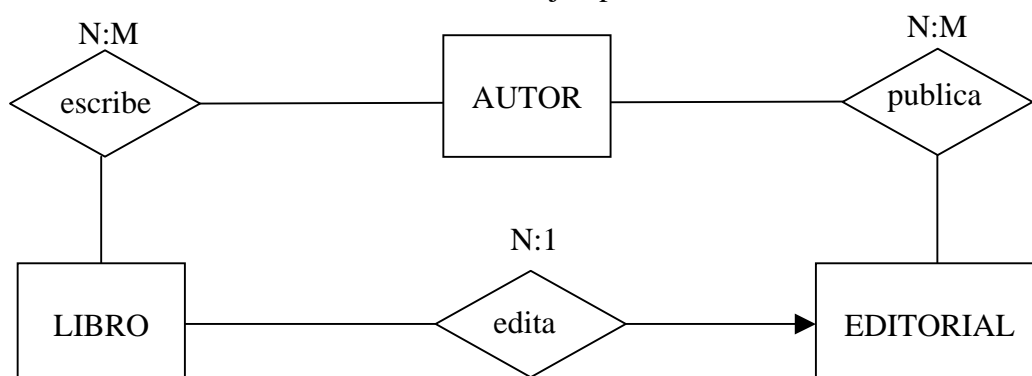
Por último, queda comentar algunas características o propiedades de los atributos que se pueden utilizar en la representación del modelo E/R o que, al menos, debemos tener en cuenta.

- **Atributos simples y compuestos.** Los atributos simples son aquellos que hacen referencia a una información que no se puede dividir en subpartes (en otros atributos) y los compuestos, los que sí. Por ejemplo, cuando especificamos una fecha nos referimos al día, mes y año, pero luego nos puede interesar obtener información por ejemplo teniendo en cuenta sólo el año. Estos atributos se deberán tratar como compuestos cuando los implementemos en la base de datos.
- **Atributos univalorados y multivalorados.** Los atributos univalorados son aquellos que únicamente van a tomar un valor, mientras que los multivalorados van a indicarnos la posibilidad de tener más valores, por lo que habrá que especificar el número mínimo y máximo de valores a tomar. Por ejemplo, en el caso de un cliente de un banco, podríamos tener los atributos *teléfono_casa*, *teléfono_móvil* y *teléfono_trabajo*, que representarían tres atributos univalorados, pero a su vez, podríamos haber considerado un único atributo multivalorado llamado *teléfono* donde le indicáramos que como máximo podemos introducir hasta tres teléfonos. El uso de atributos multivalorados en la representación puede ser útil siempre que no olvidemos que en realidad no es más que un número determinado de atributos diferentes. La forma de representar un atributo multivalorado es mediante una doble elipse en lugar de una simple.
- **Atributo nulo.** Consideraremos que un atributo puede ser nulo cuando la entidad no tiene un valor para ese atributo.
- **Atributo derivado.** Un atributo es derivado cuando su valor se puede obtener a partir de otros atributos o entidades. La forma de representarlo es mediante una elipse discontinua. Este tipo de atributos se deben omitir dado que daría redundancia de información.

2.2. Control de redundancia de los diagramas E/R

Una vez construido un esquema E/R, hay que analizar si se presentan redundancias, ya que estas pueden acarrear problemas a la hora de construir la base de datos real. Además de la existencia de atributos redundantes, como los que se derivan de otros mediante algún cálculo, hay que estudiar detenidamente los ciclos en el diagrama E/R, ya que pueden indicar la existencia de relaciones redundantes.

Un ciclo, en un diagrama E/R, es un grupo de tipos de entidad unidos en forma circular o cíclica a través de ciertas relaciones. Por ejemplo:



Es evidente que en este ejemplo podríamos eliminar la relación *publica*, puesto que si conocemos los libros que un autor ha escrito, y sabemos las editoriales que editan dichos libros, somos capaces de extraer las editoriales en las que un determinado autor publica sus libros. Por consiguiente, podemos eliminar la relación publica de nuestro diagrama por ser redundante.

Este caso no es general, es decir, es posible que existan ciclos en los que no aparezcan redundancias. Los casos más típicos son los que resultan de dividir relaciones ternarias en sus correspondientes binarias.

3. Modelo entidad relación extendido (EE/R)

Una vez visto el modelo básico de Chen, hay que profundizar un poco más en otros aspectos que aportan aún más significado y relevancia a esta herramienta.

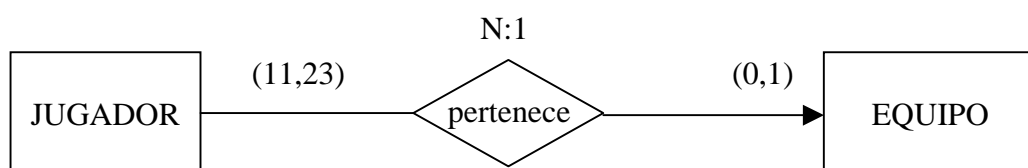
Ya hemos señalado que varios autores han considerado diversas extensiones al modelo E/R definido por Chen, dando lugar a lo que algunos han llamado modelo E/R extendido (EE/R). Trataremos de mostrar las extensiones al modelo que resultan más interesantes por su funcionalidad y uso.

3.1. Cardinalidad de un tipo de entidad

Se define como el número máximo y mínimo de ocurrencias de un tipo de entidad que pueden estar relacionadas con una ocurrencia del otro u otros tipos de entidad que participan en la relación. Su representación gráfica es una etiqueta del tipo (0,1), (1,1), (0,n), o (1,n), según corresponda. El significado de cada una es:

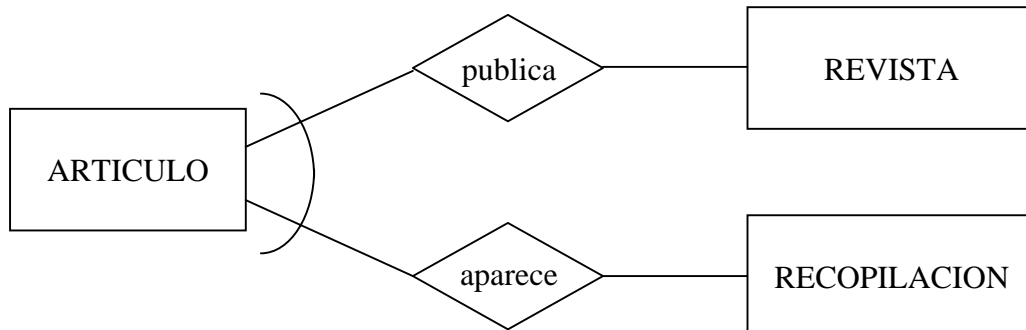
- (0,1). Para una ocurrencia determinada de los otros tipos de entidades que participan en la relación, el valor mínimo de ocurrencias de la entidad que se trata es 0, y el número máximo es 1.
- (1,1). Para una ocurrencia determinada de los otros tipos de entidades que participan en la relación, debe existir una y sólo una ocurrencia de la entidad que se trata.
- (0,n). Para una ocurrencia determinada de los otros tipos de entidades que participan en la relación, el valor mínimo de ocurrencias de la entidad que se trata es 0, y el número máximo es n .
- (1,n). Para una ocurrencia determinada de los otros tipos de entidades que participan en la relación, debe existir como mínimo una ocurrencia de la entidad que se trata, si bien no hay límite en el número de veces que dicha ocurrencia puede aparecer en la relación.

Supongamos como ejemplo el caso de un equipo de fútbol, sus cardinalidades serían:



3.2. Relaciones exclusivas

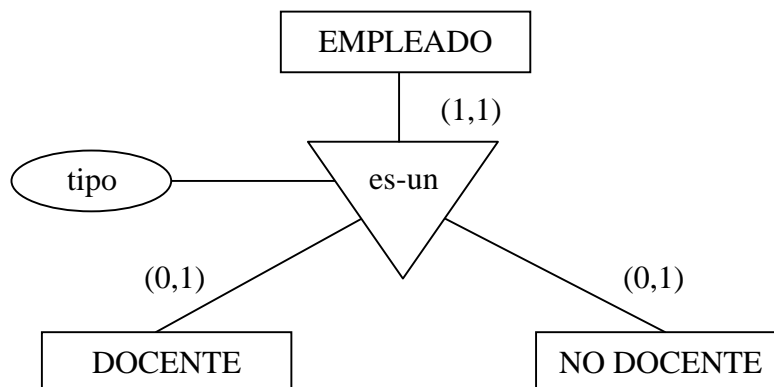
Decimos que dos o más tipos de relación son exclusivos cuando cada ocurrencia de un tipo de entidad sólo puede pertenecer a un tipo de relación. Veamos un ejemplo con su correspondiente forma de representación:



3.3. Generalización y Herencia

La descomposición de tipos de entidad en varios subtipos es una necesidad muy habitual en el modelado de bases de datos. En el mundo real se pueden identificar varias jerarquías de entidades. La relación que se establece entre un supertipo y sus subtipos corresponde a la noción de 'es_un' o 'es_un_tipo_de'.

Este tipo de relación especial se representa a través de un triángulo invertido con la base paralela al rectángulo que representa al supertipo, y conectado a los subtipos. Esta relación tiene la característica de que toda ocurrencia de un subtipo es una ocurrencia del supertipo, aunque no sucede lo contrario, con lo que las cardinalidades serán siempre (1,1) en el supertipo, y (0,1) o (1,1) en el subtipo. El atributo del supertipo que actúa como discriminante se liga al triángulo a través de una elipse. Por ejemplo:

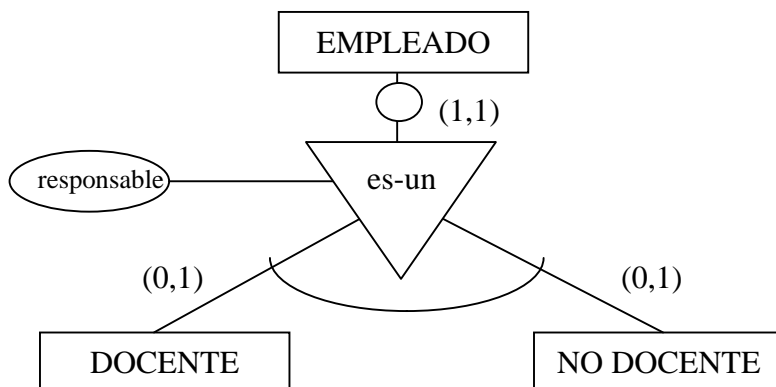


Una característica importante en estas relaciones es la herencia, puesto que cualquier atributo del supertipo pasa a ser un atributo de los subtipos.

Adicionalmente, existen dos tipos de elementos que se utilizan en la representación para indicar lo siguiente:

- *Arco*. Indica la exclusividad de los subtipos, es decir, que una entidad del supertipo sólo puede ser de uno sólo de los subtipos.
- *Elipse vacía*. Indica la obligatoriedad del supertipo de pertenecer a alguno de los subtipos.

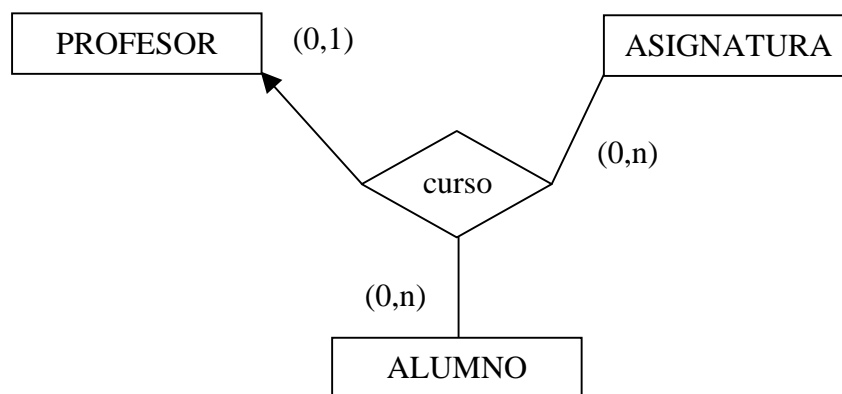
Un ejemplo completo de estos diagramas sería el mostrado en la siguiente figura, en la que se indica que un empleado debe ser de tipo docente o no docente, y además no puede pertenecer a los dos subtipos a la vez.



3.4. Relaciones Ternarias (de grado 3)

Como ya se ha mencionado con anterioridad, lo habitual es encontrar relaciones binarias en los diagramas E/R. Sin embargo, esto no significa que no puedan existir relaciones de orden superior, aunque siempre es conveniente tratar de reducirlas (en la medida en que la semántica de la relación lo permita) a relaciones binarias. Vamos a analizar cómo se representan las relaciones ternarias, y cómo se deben interpretar las cardinalidades de estos tipos de relaciones.

Supongamos la relación ternaria mostrada en la figura inferior. En esta relación de ejemplo aparecen las entidades Profesor, Asignatura y Alumno. La relación se llama curso, y representa las ternas que se pueden dar durante un curso académico, relacionando los profesores que imparten ciertas asignaturas, con los alumnos matriculados en dichas asignaturas, y por consiguiente, los profesores que tienen asignados ciertos alumnos.



Es evidente que esta relación se podría descomponer en tres relaciones binarias, de modo que los profesores estuviesen relacionados con los alumnos a los que les imparten clase, los alumnos estuviesen relacionados con las asignaturas en las que están matriculados, y las asignaturas estuviesen relacionadas con los profesores que las imparten.

La información que se podría extraer de ambos diagramas E/R sería la misma, sólo que quizás se encontraría de forma más compacta en la representación tabular

correspondiente al diagrama con la relación ternaria. No obstante, es mucho más fácil de interpretar el diagrama E/R con relaciones binarias.

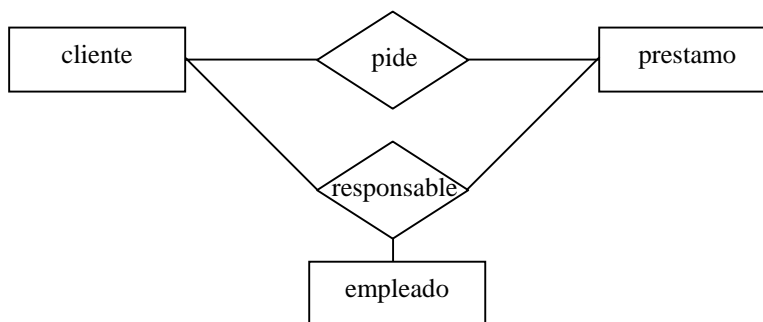
Volviendo al diagrama con la relación ternaria, hay que indicar que todos los elementos que pueden aparecer en torno a la relación tienen el mismo significado que en las relaciones binarias, excepto la cardinalidad de las relaciones.

Como podrá observarse fácilmente, la cardinalidad de una relación binaria hace referencia al número de apariciones que puedo tener (máximo y mínimo) de una entidad respecto a la aparición de un elemento de la otra entidad que participa en la relación. Si se pretende extender este significado al caso de relaciones de orden n , habrá que especificar la cardinalidad máxima y mínima de elementos de una entidad respecto a los $n-1$ elementos de las $n-1$ entidades restantes. Es decir, hay que fijar la aparición de elementos de $n-1$ entidades para encontrar la cardinalidad máxima y mínima de los elementos de la entidad restante.

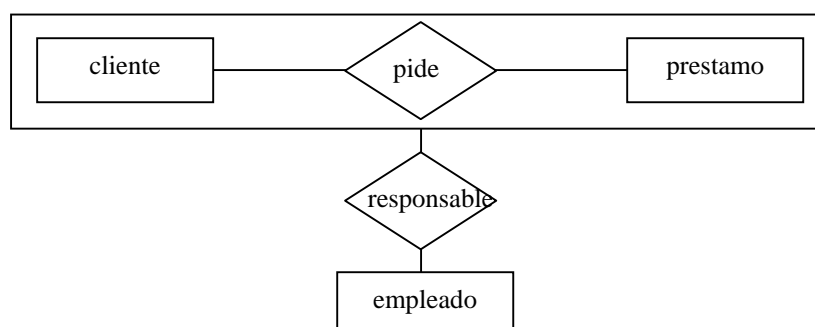
Por tanto, en nuestro ejemplo, la cardinalidad $(0,n)$ correspondiente a la entidad alumno indica que: para una pareja fija de profesor/asignatura puedo tener en la relación un mínimo de 0 alumnos y un máximo sin determinar.

3.5. Agregación

Una limitación del modelo E/R básico es que no es posible expresar relaciones entre relaciones sin que esto conlleve redundancias de información o bien, si tratamos de simplificarlo, nos obscurezca el esquema de forma que el diseño conceptual quede poco claro. Por ejemplo, si tratamos de representar la situación de un cliente de un banco que pide un préstamo y para ese par cliente-préstamo podemos tener un empleado del banco que sea el responsable de ello, la forma en que podríamos construir el modelo sería:



Pero así tenemos información redundante. La mejor forma de modelar esta situación es usando una *agregación*. La **agregación** es una abstracción a través de la cual las relaciones se tratan como entidades de nivel más alto. Así, en nuestro ejemplo, consideraríamos el tipo de relación *pide* y los tipos de entidades *cliente* y *prestamo* como si fueran un tipo entidad llamado *pide* y se trataría como si fuera cualquier otro tipo entidad. Por tanto, la forma de representar el esquema sería:



4. Ejemplo

Tomemos como ejemplo el diseñar una base de datos relacional que permita la gestión de préstamos de libros de una biblioteca y, como primer paso, vamos a desarrollar el diseño conceptual de dicha base de datos, es decir, vamos a desarrollar el modelo E/R.

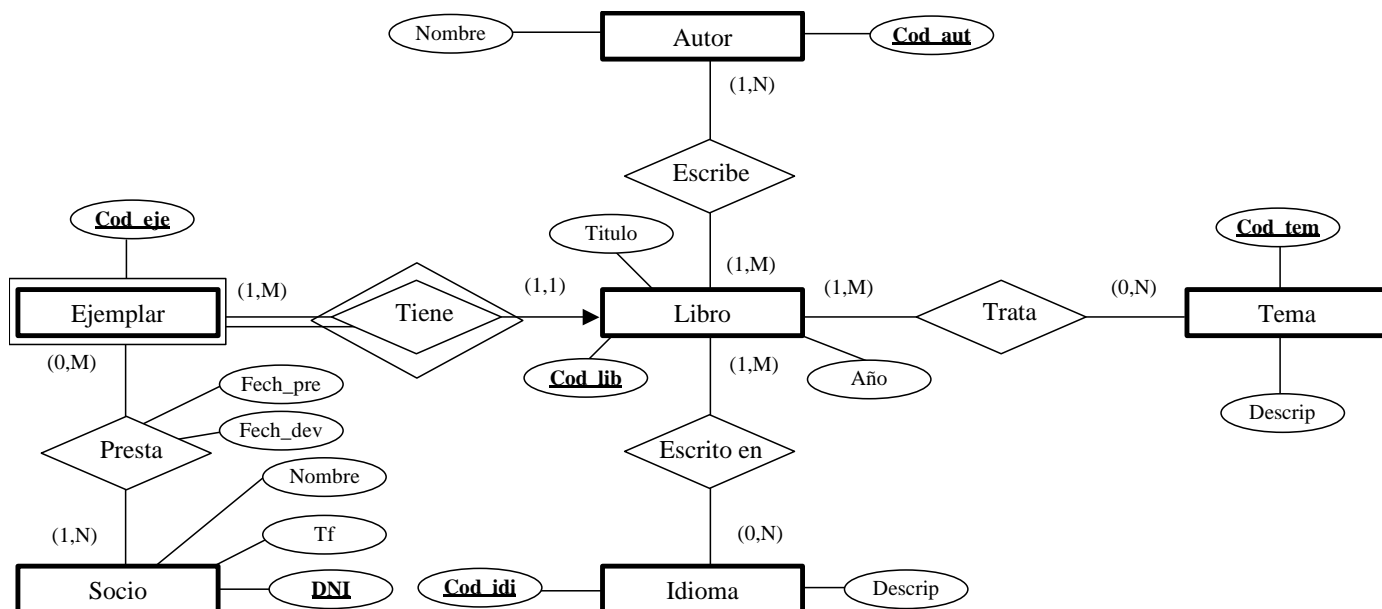
En nuestro ejemplo de estudio partimos de la forma actual de trabajo de la biblioteca en cuestión, la cual consiste en una serie de fichas de tres tipos:

- Fichas con las características de los libros (nombre, código, tipo, etc.).
- Fichas con las características de los lectores (nombre, apellidos, domicilio, etc.).
- Fichas con la información de los préstamos de libros que se han efectuado, incluyendo el lector a quién se le ha prestado, la fecha, etc.

Además de estas fichas, en nuestras conversaciones con los empleados, obtenemos algunas informaciones y comentarios útiles para el diseño que son los siguientes:

- De cada libro pueden existir varios ejemplares.
- Se está interesado en tener información sobre el/los idioma/s del libro.
- Interesa reflejar los temas de los libros, pudiendo cada libro pertenecer a varios temas y/o subtemas.
- Interesa conocer el nombre de los autores.

A partir de esta información podemos obtener el siguiente diseño conceptual, donde se incluye el tipo de correspondencia entre las entidades y su cardinalidad.



El tipo de correspondencia es obtenido en base a las posibilidades de relación entre las entidades. Así por ejemplo, un libro puede haber sido escrito por varios autores (relación 1:N), pero además, un autor puede haber escrito varios libros (relación 1:N en sentido inverso), por lo cual, la relación resultante es N:M.

Con la obtención del esquema de la figura puede considerarse finalizado el diseño conceptual, pues tenemos identificados las entidades y las relaciones entre ellas, así como sus cardinalidades, pudiendo pasar al diseño lógico que es el tema siguiente.