



CENTRO DE ESTUDIOS FINANCIEROS

VIRIATO, 52	28010 MADRID	914 44 49 20
PONZANO, 15	28010 MADRID	914 44 49 20
G. DE GRÀCIA, 171	08012 BARCELONA	934 15 09 88
ALBORAYA, 23	46010 VALENCIA	963 61 41 99

www.cef.es

info@cef.es

Índice Tema 6

Introducción.

1. La modelización de datos. El modelado conceptual.
2. Concepto de modelo de datos.
 - 2.1. Clasificación de los modelos de datos.
3. El Modelo Entidad/Relación.
 - 3.1. Elementos del Modelo Entidad/Relación.
 - 3.1.1. Entidad.
 - 3.1.2. Asociación, Relación o Interrelación.
 - 3.1.3. Atributo.
 - 3.1.4. Restricciones o limitaciones de integridad.
 - 3.2. Etapas del modelo conceptual de datos.
 - 3.3. Ejemplo de aplicación de la técnica del Modelo E/R.
4. El Modelo Entidad/Relación Extendido.
 - 4.1. Un ejemplo de Modelo Entidad/Relación Extendido.



CENTRO DE ESTUDIOS FINANCIEROS

VIRIATO, 52	28010 MADRID	914 44 49 20
PONZANO, 15	28010 MADRID	914 44 49 20
G. DE GRÀCIA, 171	08012 BARCELONA	934 15 09 88
ALBORAYA, 23	46010 VALENCIA	963 61 41 99

www.cef.es

info@cef.es

TEMA 6

Modelización conceptual. Elementos: entidades, atributos, relaciones, identificadores, roles. Reglas de modelización. Validación y construcción de modelos de datos.

INTRODUCCIÓN.

El desarrollo tradicional o estructurado de un Sistema de Información, a diferencia del desarrollo orientado a objetos, mantiene una clara separación entre los datos y las funciones o procesos del sistema.

La Modelización de Datos es una actividad que se realiza a lo largo del proceso de desarrollo del Sistema de Información en dos momentos diferentes:

1. En el proceso de análisis, en concreto en el análisis de datos, en que se lleva a cabo el modelado conceptual de datos.
2. En el proceso de diseño, en el que se lleva a cabo el modelado lógico de datos durante el diseño de datos, y el modelado físico, durante el diseño físico de datos.

Las ventajas de la modelización de datos se derivan de la noción de independencia. Cualquier cambio en la información, en los requisitos del usuario, en el sistema de gestión de la base de datos o en los equipos físicos, se podrá afrontar de forma mucho más eficaz e igualmente, el mantenimiento posterior del sistema se podrá realizar de forma más sencilla.

En este tema nos centraremos exclusivamente en el proceso de análisis y, por consiguiente, sólo nos referiremos a la modelización conceptual de datos.

El análisis, siguiendo la aproximación clásica, consta de tres tareas:

- Educación, cuyo objetivo es adquirir el conocimiento del dominio del problema, de forma que sea posible identificar los requisitos del sistema.
- Modelización, cuyo objetivo es representar, mediante la utilización de «Modelos Conceptuales», los conocimientos adquiridos en la tarea anterior.
- Validación, cuyo objetivo es verificar la exactitud de los conocimientos adquiridos.

Los modelos conceptuales son mecanismos de representación que permiten registrar los requisitos educidos con el fin de facilitar su comprensión y permitir su comunicación entre todos los participantes en la actividad de análisis (clientes, usuarios y analistas). Un modelo conceptual está formado por dos elementos: un procedimiento, que permite al analista estudiar el problema, y un formalismo de representación gráfica con el que manipular los distintos componentes del problema y documentar los resultados.

Durante el análisis se investigan todos los aspectos del dominio del problema, y para ello, los elementos a considerar para describir el comportamiento externo del software, son:

- Los objetos, entendiendo por tal, una entidad del mundo real claramente delimitada, perteneciente al dominio del problema.
- Las funciones, es decir, las tareas o actividades que tienen que ser ejecutadas por el sistema que se está especificando, para resolver el problema.
- Los estados, esto es, las condiciones del sistema, objeto o función, que ayudan a capturar su historia, de forma que define cómo se comporta en condiciones específicas.

Esto quiere decir que para realizar la actividad de análisis es necesario definir varios modelos conceptuales, ya que no existe un único modelo que permita representar todos los aspectos del problema a la vez. Según esto, el problema debe modelizarse atendiendo a tres descripciones, aunque distintas, complementarias.

- La descripción estática identifica la estructura estática de los objetos y sus relaciones. Se representa mediante el Modelo de Datos (enfoque estructurado) o el Modelo de Objetos (enfoque orientado a objetos).
- La descripción funcional describe cómo se transforman los datos en el contexto del problema. Se representa mediante el Modelo de Procesos.
- La descripción dinámica describe los aspectos del sistema que cambian con el tiempo y que caracterizan los distintos estados que puede presentar el problema, así como las condiciones y eventos que permiten transitar entre estos estados. Se representa mediante el Modelo de Estados.

Asimismo, para realizar la actividad de análisis el analista dispone de un conjunto de métodos y técnicas específicas que proporcionan una serie de mecanismos que permiten representar el dominio del problema. Estos métodos y las técnicas específicas responden a cuatro enfoques diferentes:

- Enfoque orientado a Datos.
- Enfoque orientado a Objetos.
- Enfoque orientado a Procesos.
- Enfoque orientado a Estados.

Cada uno de estos enfoques explota un conjunto distinto de conceptos del dominio del problema, en el que centra su atención. La diferencia entre los distintos enfoques está definida por el hecho de cada uno prima un conjunto de conceptos sobre los restantes y considera que dichos conceptos son los únicos relevantes para el desarrollo posterior.

En el enfoque orientado a datos, objeto de este tema, el concepto fundamental a identificar son los datos o elementos de información y las relaciones existentes entre ellos. El analista considera que

los datos que puede identificar y aislar representan el estado del dominio del problema desde el punto de vista estático, y por consiguiente, los métodos asociados a este enfoque representan la visión estática del dominio de información del problema y permiten identificar la estructura estática de las entidades de datos y las relaciones existentes entre ellas, es decir, el Modelo Conceptual de Datos. La técnica más popular del enfoque orientado a datos es el modelo E/R (Entidad/Relación).

Así pues, para lograr el objetivo fundamental de este tema, que no es otro que la obtención, representación y comprensión del Modelo Conceptual de Datos, lo primero debe ser comprender qué significa la modelización de datos y qué ventajas reporta ésta tanto al análisis como al diseño del sistema, y en qué consiste el modelado conceptual.

El segundo objetivo de este tema es la representación del modelo conceptual de datos según la técnica del Modelo Entidad / Relación. A estos efectos, se estudiará exhaustivamente este modelo, detallando los elementos que lo componen y exponiendo las etapas a seguir para la obtención del modelo conceptual de datos. Todo ello se pondrá de manifiesto mediante un ejemplo práctico. Asimismo, se hará referencia al Modelo Entidad / Relación Extendido, como evolución del modelo de Chen.

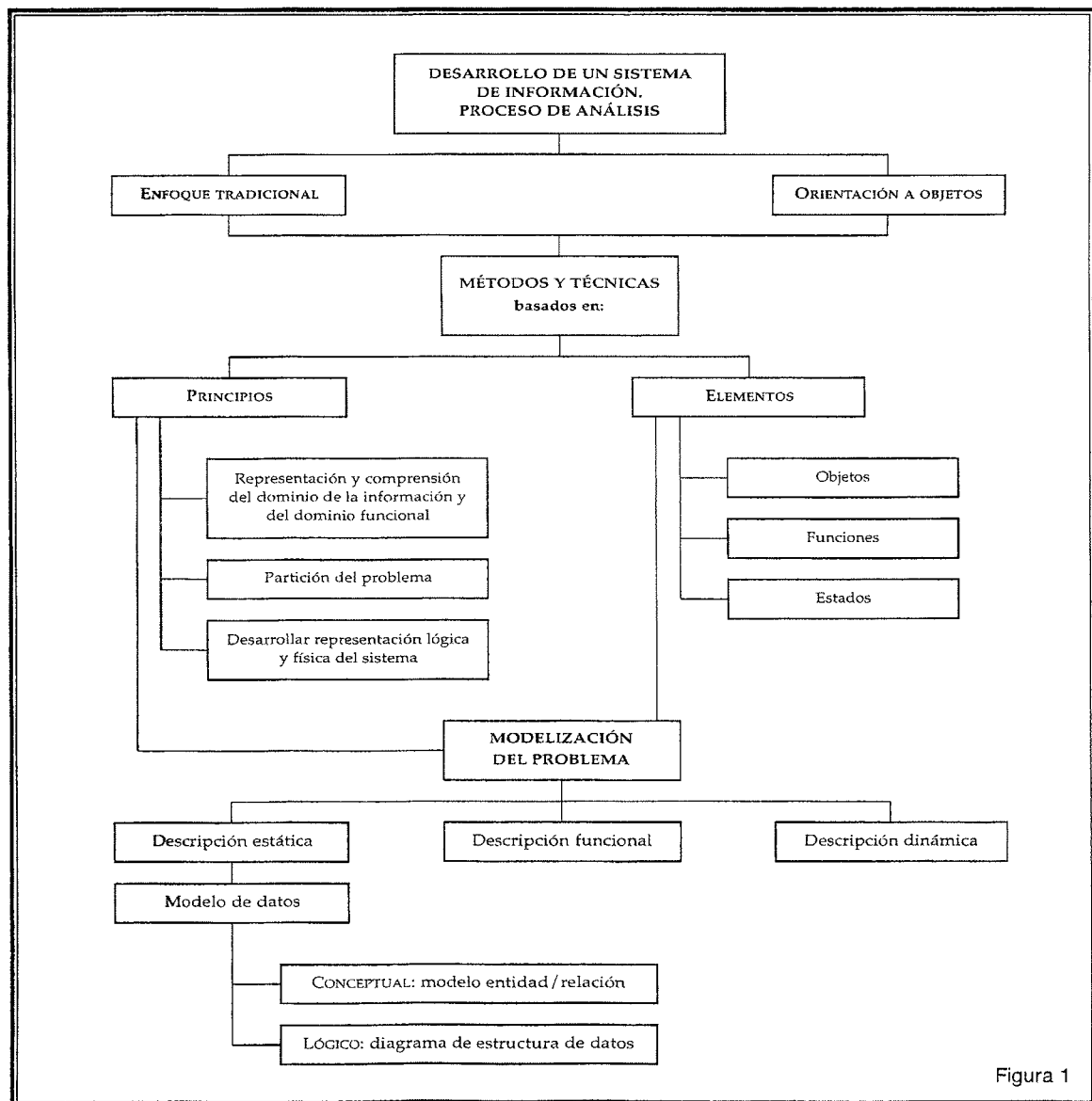


Figura 1

1. LA MODELIZACIÓN DE DATOS. EL MODELADO CONCEPTUAL.

Como ya se ha indicado, la modelización de datos es una actividad que se lleva a cabo tanto durante el proceso de análisis (modelización conceptual) como durante el proceso de diseño (modelización lógica y modelización física).

Partiendo de la premisa de que prácticamente la totalidad de las aplicaciones y sistemas de información se basan en la utilización de bases de datos para almacenar su información, uno de los principales objetivos del modelado de datos es conseguir la independencia entre las estructuras lógica y física de los datos puesto que conseguido este objetivo estará garantizada la independencia entre los datos y las aplicaciones, requisito imprescindible para satisfacer las necesidades de cambios, ya sea en los datos o en los requerimientos de los usuarios.

En este sentido, el Organismo de Estandarización de Estados Unidos ANSI (American National Standards Institute) propuso una arquitectura a tres niveles de abstracción en el SGBD, que proporciona el máximo grado de independencia física/lógica de los datos a nivel de descripción.

Los tres niveles que propone la arquitectura ANSI son:

1. Nivel Conceptual, que está orientado hacia la visión lógica del conjunto de información que proviene del mundo real, y se pretende reflejar la estructura y las relaciones existentes entre los datos del mundo real que se van a almacenar en la base de datos (universo del discurso), aislando entre sí el nivel externo (vista del usuario) y el nivel físico (vista de la máquina). Este nivel es el que se corresponde con la modelización conceptual de datos.
2. Nivel Lógico global o Externo, que está orientado hacia el usuario y comprende las distintas vistas parciales de la base de datos que se muestran a los usuarios, esto es, las características lógicas de los datos para los programas de aplicación. Es el nivel correspondiente a la modelización lógica.
3. Nivel Físico o Interno, que está orientado hacia la máquina y comprende las características de tipo físico, es decir, qué, cómo y dónde se van a almacenar los datos físicamente. Es el nivel correspondiente a la modelización física.

Por consiguiente, el primer objetivo de la modelización de datos es obtener una representación de la información que ha de manejar el sistema, independiente de las aplicaciones y de la plataforma física (hardware) en que se vaya a implantar. Es decir, obtener un modelo conceptual de datos, que permita describir, de un modo totalmente independiente de la implementación, los datos que el usuario quiere recoger en el sistema (la información que ha de manejar el sistema se suele llamar «universo del discurso», entendiendo por tal, la parte o visión del mundo real que es relevante para el sistema).

El modelado conceptual de datos es una actividad que se realiza en la etapa de Análisis, paralelamente al modelado de procesos. Su objetivo es captar toda la información del mundo real que es relevante para el sistema, abstrayendo los detalles sin importancia, y sin considerar cómo se va a implementar el sistema. De hecho, dependiendo del tipo de sistema, del volumen de información, de los requisitos de eficiencia, etc., se podrán utilizar distintos mecanismos de persistencia de los datos (Sistemas de Bases de Datos, Sistemas de Ficheros, etc.), pero esto corresponde a la etapa de Diseño.

La siguiente figura muestra la situación del modelado conceptual de datos en el proceso de desarrollo, así como la relación existente entre ésta y otras actividades.

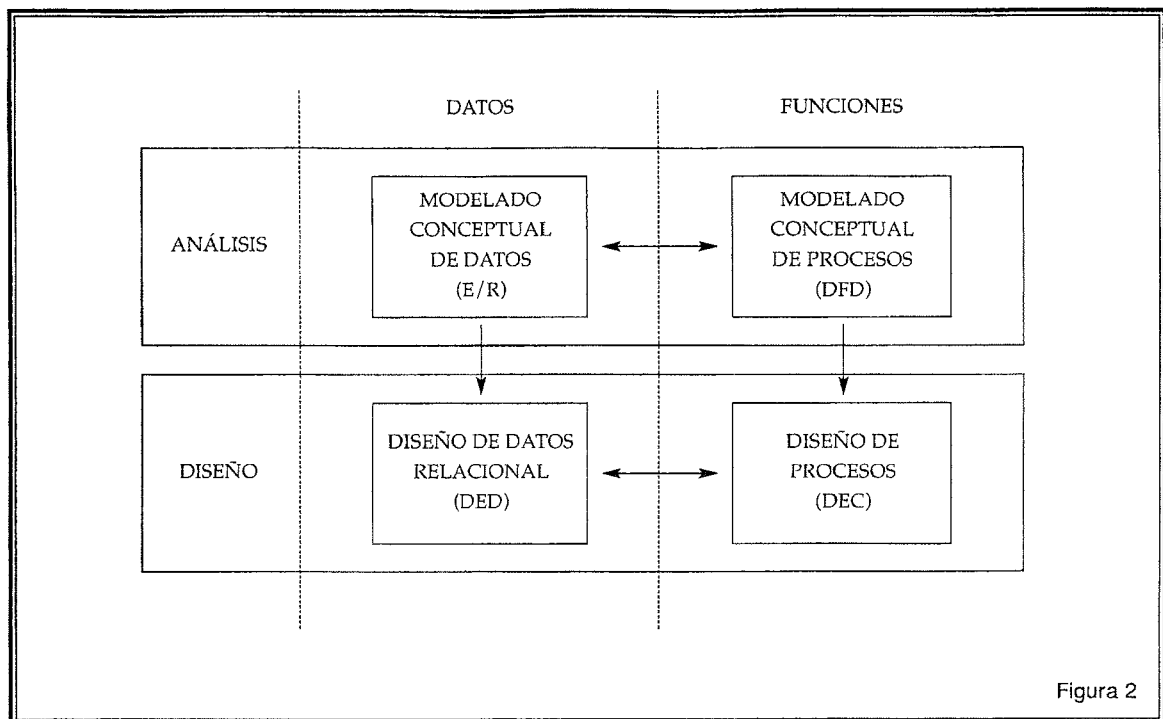


Figura 2

Durante el Análisis lo que interesa es recoger la máxima cantidad de información posible descomponiendo el universo del discurso del sistema en sus componentes para estudiarlos tanto de manera aislada, como en interacción con el resto, a fin de conseguir una modelización óptima de la información que maneja el sistema; actividad que es crítica para asegurar el éxito del sistema de información.

Para ello, se ha de obtener la percepción que tiene el usuario del universo del discurso, esto es, el «real percibido» por el usuario y representarlo, utilizando alguna técnica formal que sea independiente de los modelos de implementación, que tenga una capacidad semántica alta y que sea lo más cercana posible al usuario. Esta representación constituye el modelo conceptual de datos.

Aunque existen diversas técnicas, utilizaremos el modelo E/R porque además de cumplir los requisitos anteriores, es la técnica de modelado conceptual de datos universalmente aceptada para el desarrollo estructurado.

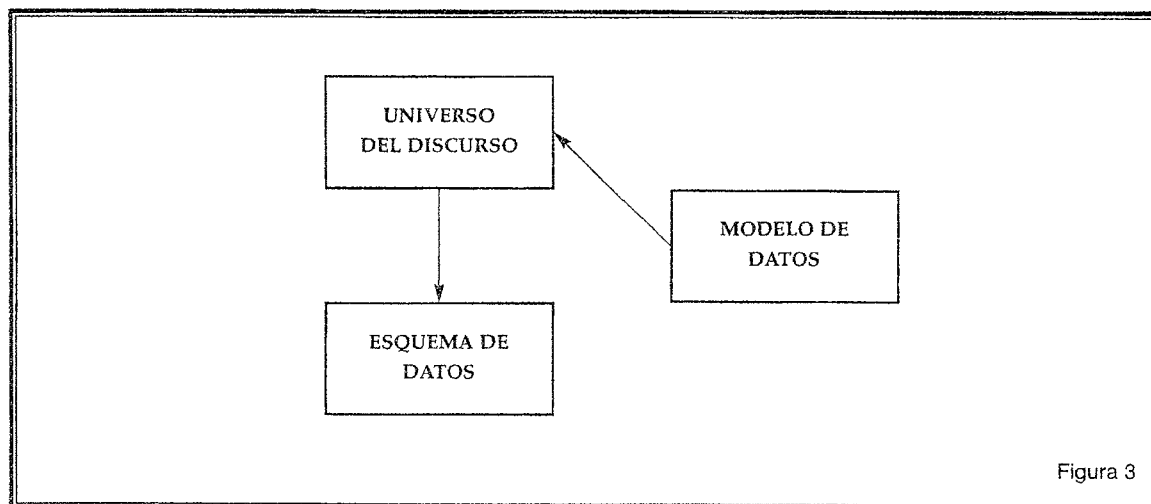
El Modelo Conceptual de Datos representa la visión estática del dominio de la información del problema, y permite identificar la estructura estática de las entidades de datos y las relaciones existentes entre ellas. En definitiva, es el «instrumento» que facilita la representación de las necesidades del usuario y debe ser una fiel representación del sistema de información objeto de estudio. Por todo ello:

- El contenido del modelo de datos debe representar el estado final al que quiere llegar el sistema.
- Cualquier cambio en el sistema de información se debe reflejar en el modelo de datos y viceversa.
- En el modelo de datos debe aparecer representada toda la información que ha de manejar el sistema (universo del discurso).

2. CONCEPTO DE MODELO DE DATOS.

Sin entrar en las distintas acepciones del término «modelo» (entendido como una reproducción simplificada de la realidad, o como la realidad propiamente dicha) y en la distinta terminología utilizada (algunos autores llaman «formalismos» a los modelos y «modelos» a los esquemas) siguiendo la terminología más extendida en el mundo de los datos, vamos a definir el Modelo de Datos como el conjunto de conceptos, reglas y convenciones que permiten describir y manipular los datos de la parcela del mundo real que constituye nuestro universo del discurso. Dicho en otros términos, el modelo de datos es un mecanismo de abstracción para interpretar la realidad, con el objetivo de captar su semántica.

Al aplicar el Modelo de Datos, es decir, al aplicar este conjunto de conceptos, reglas y convenciones, a un cierto universo del discurso se obtiene una estructura de datos llamada Esquema. Según esto, definimos el Esquema de Datos como la representación de un determinado universo del discurso en términos de un modelo de datos.



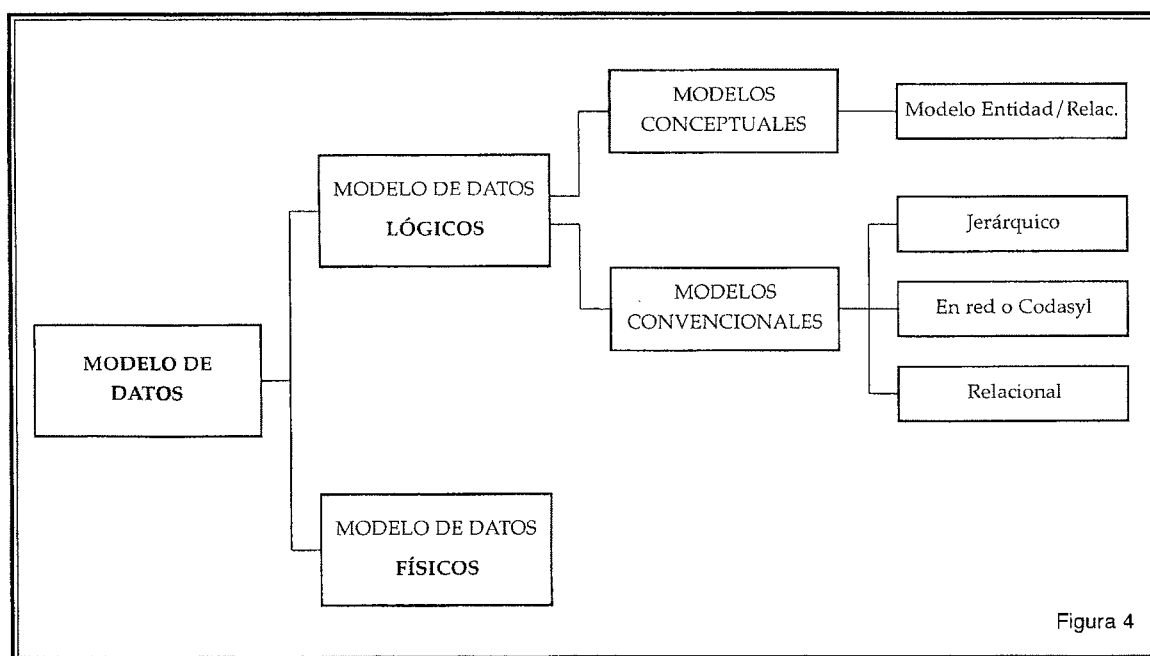
Existen distintos modelos de datos para construir los esquemas correspondientes a cada nivel de abstracción de la arquitectura ANSI:

- El esquema conceptual, que debe captar y almacenar el «universo del discurso» que ha de tratar el sistema de información, y sirve de punto de control para futuros desarrollos de la base de datos, aísla la representación de la información de los requerimientos de la máquina y de las exigencias de cada usuario, e independiza la definición de la información de los SGBD. En este esquema, al analista sólo le concierne el aspecto global de la organización o unidad que se trata, y en él se definen las entidades de datos y las relaciones entre ellas en un nivel de abstracción próximo al usuario. Es el correspondiente al modelo de datos conceptual.
- El esquema lógico o externo, que permite «ver» a cada tipo de usuario de la base de datos, sólo aquella parte que es de su interés. Al haber diferentes usuarios con distintas necesidades, existirán distintos esquemas lógicos o externos para un mismo esquema conceptual.
- El esquema físico o interno. Es el que especifica cómo son almacenados los datos en memoria, describe la estructura de la base de datos en forma de modelo conceptual de almacenamiento y depende del entorno físico donde se vaya a implantar el sistema.

2.1. CLASIFICACIÓN DE LOS MODELOS DE DATOS.

Tampoco hay consenso a la hora de clasificar los distintos tipos de modelos de datos. Mientras algunos autores distinguen dos tipos de modelos (lógicos y físicos) y dentro de los primeros hacen la distinción entre modelos convencionales y modelos conceptuales, tal como se muestra en la siguiente figura, otros, siguiendo las directrices de la Arquitectura ANSI a tres niveles y en consonancia con los esquemas definidos, consideraran la siguiente clasificación:

- Modelo de datos conceptual. Es único para cada sistema de información y se representará mediante el Modelo Entidad/Relación.
- Modelo de datos lógico. Depende del SGBD que se vaya a utilizar y por tanto no es único. El modelo lógico de datos podrá ser un modelo jerárquico, un modelo en red o un modelo relacional, si bien, dada la obsolescencia de los dos primeros SGBD, siempre se considerará que el modelo lógico de datos es el modelo relacional. Se representará mediante la técnica del Diagrama de Estructura de Datos.
- Modelo de datos físico. Depende de las características propias de la máquina (sistema operativo, sistema de gestión de ficheros, etc.) en que se vaya a implantar el sistema, y corresponde al nivel interno de la arquitectura ANSI.



3. EL MODELO ENTIDAD/RELACIÓN.

El Modelo Entidad/Relación, en adelante Modelo E/R, fue propuesto por Peter Chen en 1976, en su artículo «The Entity/Relationship Model: Toward a unified view of data». Algunos autores proponen llamarlo Modelo Entidad/Interrelación, con el fin de no confundirlo con la traducción del término inglés «relation» utilizado en el modelo relacional de Codd, o Modelo Entidad/Asociación.

El Modelo E/R es un formalismo conceptual que se puede definir como un conjunto de conceptos, reglas y convenciones que permiten describir con un alto grado de formalismo un modelo externo para convertirlo en un modelo conceptual de datos.

El Modelo E/R permite establecer una visión global de los datos de una organización o de un sistema de información en un nivel de abstracción próximo al usuario e independiente de las características del equipo donde después se vaya a instrumentar el sistema. El modelo consiste en describir la información de la organización mediante la definición de entidades de datos y asociaciones o relaciones entre las mismas.

Este modelo se corresponde con el nivel conceptual de la Arquitectura ANSI, por lo que también se le conoce, de forma genérica, como «Modelo Conceptual de Datos».

3.1. ELEMENTOS DEL MODELO ENTIDAD/RELACIÓN.

Los elementos fundamentales de que consta el Modelo E/R básico son los siguientes:

- Entidades.
- Asociaciones, relaciones o interrelaciones entre las entidades.
- Atributos.
- Restricciones o limitaciones de integridad.

3.1.1. Entidad.

Se define entidad como cualquier objeto real o abstracto, que tiene existencia por sí mismo, se puede identificar de una manera clara y precisa y sobre el que se quiere tener información en el sistema. Por ejemplo, si nuestro universo del discurso es una empresa, «Empleado» sería una entidad puesto que parece lógico que se quiera tener información al respecto en el sistema.

Toda entidad, al igual que el resto de componentes del modelo E/R debe quedar unívocamente identificada por un nombre, que debe ser único en el modelo y lo más significativo posible. Normalmente las entidades se nombran con un sustantivo.

Se denomina ocurrencia o instancia a cada uno de los elementos con características comunes que componen una entidad. Por ejemplo, en la entidad «Empleado», los empleados Pedro, Juan, etc., son ocurrencias de la misma.

En su teoría, Peter Chen llama «entity-type», esto es, «Entidad-Tipo» a cada clase de objetos perteneciente al universo del discurso, es decir, a la estructura genérica. Es lo que nosotros aquí hemos denominado entidad. De la misma forma, Chen llama «entity», o sea, «Entidad» a cada una de las ocurrencias o instancias de una entidad-tipo. No obstante, aquí mantendremos los términos «entidad» y «ocurrencia».

Utilizando un formalismo matemático, una entidad-tipo se puede definir como un conjunto de instancias que cumplen el predicado asociado a la entidad-tipo:

$$E = \{e/P(e)\}$$

siendo: E: Entidad-tipo

e: Instancia u ocurrencia de la entidad-tipo.

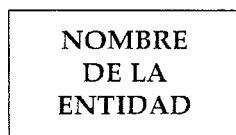
P(e): Predicado asociado a la entidad-tipo.

Dentro del modelo E/R se pueden distinguir dos clases de entidades:

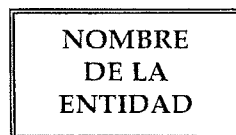
- Las entidades regulares, que son aquellas que tienen existencia propia. Por ejemplo, en el universo del discurso de una empresa, la entidad «Empleado».
- Las entidades débiles, que son aquellas cuya existencia depende de otra entidad. Por ejemplo, si se estima conveniente almacenar información sobre los hijos que tiene un empleado, aparece la entidad «Hijo», cuya existencia en el universo del discurso de la empresa no tiene sentido si no es por la existencia de la entidad «Empleado».

Gráficamente las entidades regulares se representan mediante un rectángulo con el nombre identificativo de la entidad dentro de él, y las débiles con un doble rectángulo.

Entidad regular:



Entidad débil:



3.1.2. Asociación, Relación o Interrelación.

Se denomina asociación, relación o interrelación a la conexión o correspondencia semántica entre entidades.

Al igual que con las entidades, aquí también cabe hablar de Asociación o Relación-Tipo, como la conexión semántica entre dos o más entidades-tipo. Por ejemplo, en el universo del discurso de una empresa, entre las entidades-tipo «Empleado» y «Departamento» se puede identificar la asociación-tipo «Pertenece». Una ocurrencia o instancia de la asociación-tipo es la relación de pertenencia de una instancia de «Empleado» a una instancia de «Departamento».

Matemáticamente se puede formalizar el concepto de asociación-tipo como un conjunto de ocurrencias o instancias de asociaciones entre instancias de entidades-tipo:

$$R = \{ \langle e_1, e_2, \dots, e_n \rangle / e_1 \in E_1, e_2 \in E_2, \dots, e_n \in E_n \}$$

siendo: R: Asociación-tipo

$\langle e_1, e_2, \dots, e_n \rangle$: Instancia de la asociación-tipo.

E_i : Entidad-tipo que participa en la asociación-tipo.

n: Grado de la asociación-tipo.

Al igual que sucede con las entidades, las asociaciones se pueden clasificar en:

- Relaciones o Asociaciones regulares, si asocian dos entidades regulares.
- Relaciones o Asociaciones débiles, si asocian una entidad débil con una entidad regular.

Como los demás componentes del modelo, toda asociación o relación debe ser identificada con un nombre único que la distinga inequívocamente del resto y sea suficientemente explicativo. Normalmente las asociaciones se nombran con un tiempo verbal.

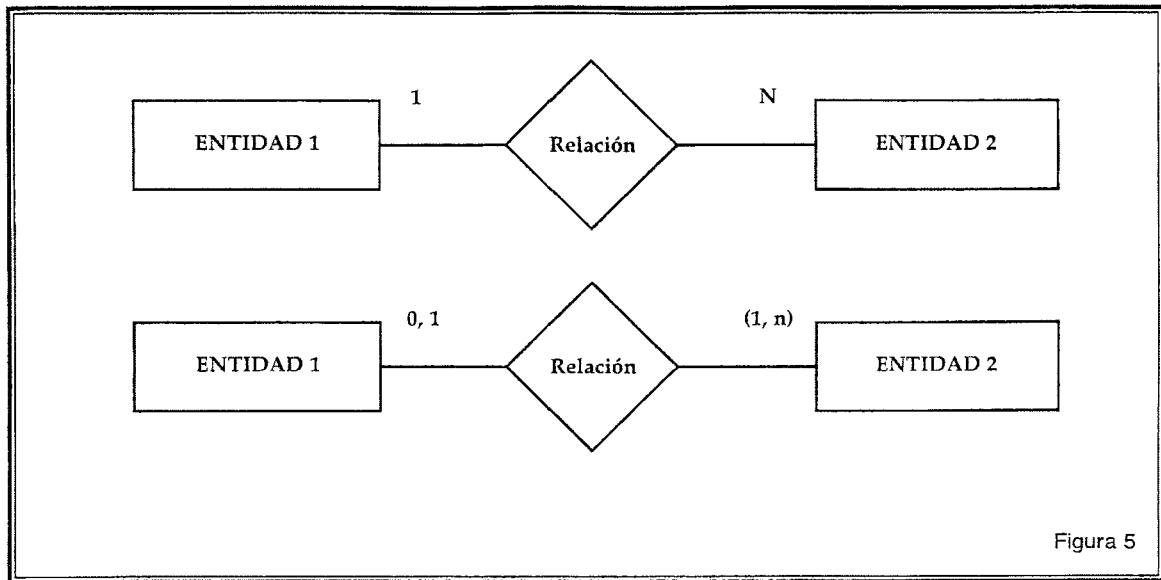
Se denomina grado de la asociación o de la relación al número de entidades que participan en una asociación dada. Atendiendo al grado las relaciones pueden ser unitarias o reflexivas (entre una misma entidad), binarias (entre dos entidades) o N-arias (entre N entidades).

Se denomina cardinalidad máxima de la asociación o tipo de correspondencia al número máximo de ocurrencias de cada entidad que pueden intervenir en la asociación o relación que se está considerando. Por este concepto las asociaciones pueden ser:

- 1: 1, si una ocurrencia de una entidad se interrelaciona como máximo con una ocurrencia de la otra entidad. Por ejemplo, dadas las entidades «Hombre» y «Mujer», la relación «está casado» es 1:1 ya que un hombre está casado como máximo con una mujer y una mujer está casada como máximo con un hombre.
- 1: N, si una ocurrencia de una entidad se interrelaciona con varias ocurrencias de la otra. Por ejemplo, dadas las entidades «Empleado» y «Departamento», la relación «pertenece» es del tipo 1: N ya que un empleado pertenece a un departamento y a un departamento pueden pertenecer varios empleados.
- N: N, si varias ocurrencias de una entidad se interrelacionan con varias ocurrencias de la otra. Por ejemplo, dadas las entidades «Empleado» y «Proyecto», la relación «trabaja» es N: N ya que un empleado puede trabajar en muchos proyectos y en un proyecto pueden trabajar muchos empleados.

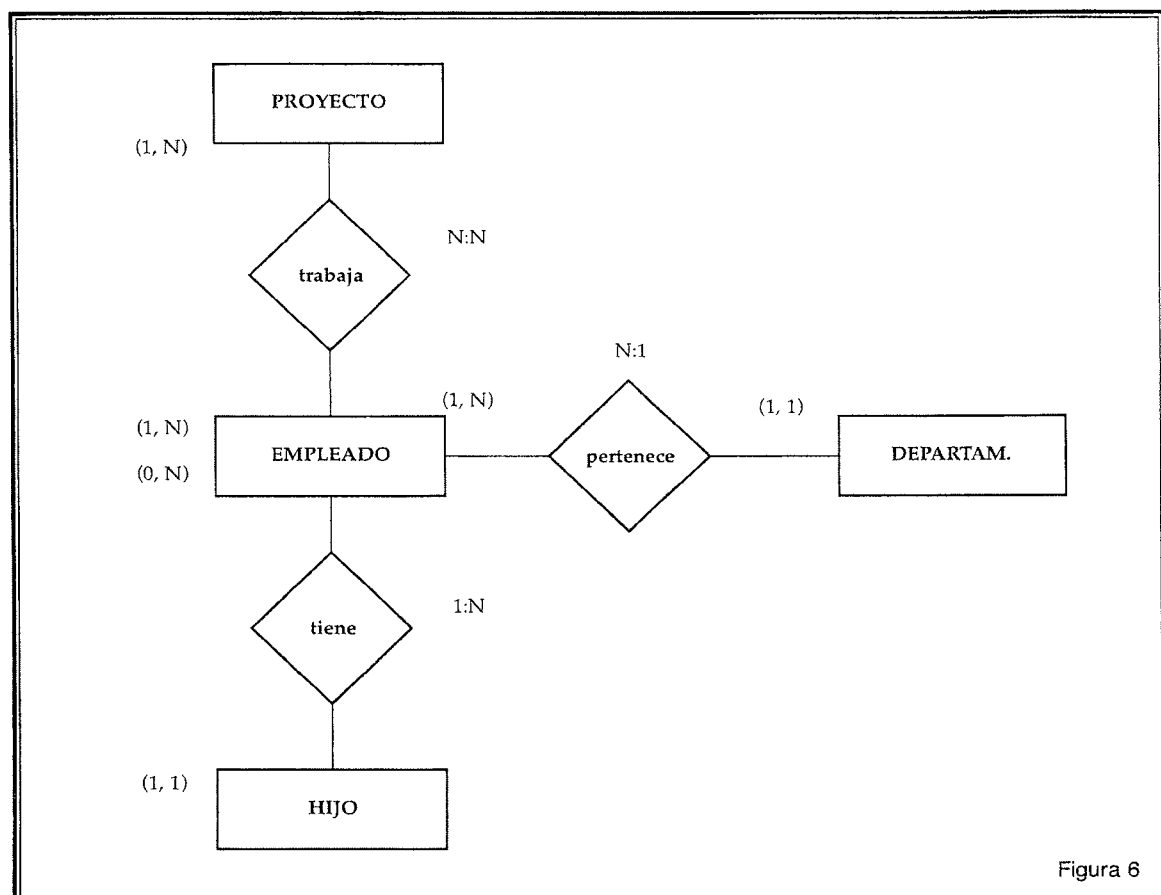
Por último, se denomina cardinalidad de una entidad al número mínimo y máximo de ocurrencias de una entidad que pueden estar interrelacionadas o asociadas con una ocurrencia de otra u otras entidades participantes en la asociación. Según corresponda, la cardinalidad de una entidad podrá ser: (0,1), (1,1), (0,N) o (1,N).

Las interrelaciones o asociaciones en el modelo E/R de Peter Chen se representan gráficamente mediante un rombo con el nombre de la interrelación o asociación dentro del mismo. Sobre el rombo figurará la cardinalidad máxima o tipo de correspondencia de la asociación y al lado de cada entidad participante en la asociación, su cardinalidad.



A título de ejemplo vamos a representar el siguiente caso:

En el universo del discurso de una empresa, cada empleado sólo pertenece a un departamento pero puede trabajar en varios proyectos. Asimismo, interesa que el sistema guarde información de los hijos que tiene cada empleado.



En efecto, respecto a la cardinalidad máxima o tipo de correspondencia de las asociaciones se cumple que:

- Un empleado pertenece a un departamento, pero a un departamento le pertenecen varios empleados, por tanto, tipo de correspondencia 1:N.
- Un empleado puede trabajar en varios proyectos y en un proyecto pueden trabajar varios empleados, por tanto, tipo de correspondencia N:N.
- Un empleado puede tener varios hijos, pero un hijo sólo lo puede ser de un empleado, por tanto, tipo de correspondencia 1, N. Al ser «Hijo» una entidad débil, esta es una asociación débil.

Respecto a la cardinalidad de las entidades se cumple:

- Asociación «Empleado» – «Departamento»: el número mínimo de empleados que pueden pertenecer a un departamento es uno, y el máximo, N. Por tanto, la cardinalidad de «Empleado» en esta asociación es (1, N). Análogamente, para la entidad «Departamento», el número mínimo y máximo de departamentos a los que puede pertenecer un empleado es uno. Por tanto, cardinalidad (1, 1).
- Asociación «Empleado» – «Proyecto»: el número mínimo de empleados que pueden trabajar en un proyecto es uno, y el máximo, N. Por tanto, la cardinalidad de «Empleado» en esta asociación es (1, N). Análogamente, para la entidad «Proyecto», el número mínimo de proyectos en los que puede trabajar un empleado es uno, y el máximo es N. Por tanto, cardinalidad (1, N).
- Asociación «Empleado» – «Hijo»: el número mínimo de empleados que pueden tener hijos es cero, y el máximo, N. Por tanto, la cardinalidad de «Empleado» en esta asociación es (0, N). Análogamente, para la entidad «Hijo», el número mínimo de hijos cuyo padre es un empleado es uno, y el máximo también es uno porque un hijo no lo puede ser de más de un padre (empleado). Por tanto, cardinalidad (1, 1).

En este ejemplo se puede observar, también, que todas las asociaciones (pertenecer, trabaja, tiene) son binarias, es decir, son de grado dos.

El modelo E/R de Chen, al ser un modelo conceptual de datos, admite en su representación gráfica todo tipo de grados en las asociaciones (reflexivas, binarias, ternarias, etc.) y todo tipo de correspondencia o cardinalidad máxima de las asociaciones (1, 1), (1, N) o (N, N). Asimismo, permite el que pueda haber más de una asociación entre dos entidades.

Otro aspecto a considerar en el modelo E/R de Chen es el concepto de asociación o relación exclusiva. A este respecto, se dice que una misma entidad que participa en dos asociaciones, lo hace de forma exclusiva si una instancia de la entidad participa en una y sólo una instancia de alguna de las asociaciones. En la representación gráfica del modelo esta circunstancia se refleja con un arco que envuelve las dos asociaciones.

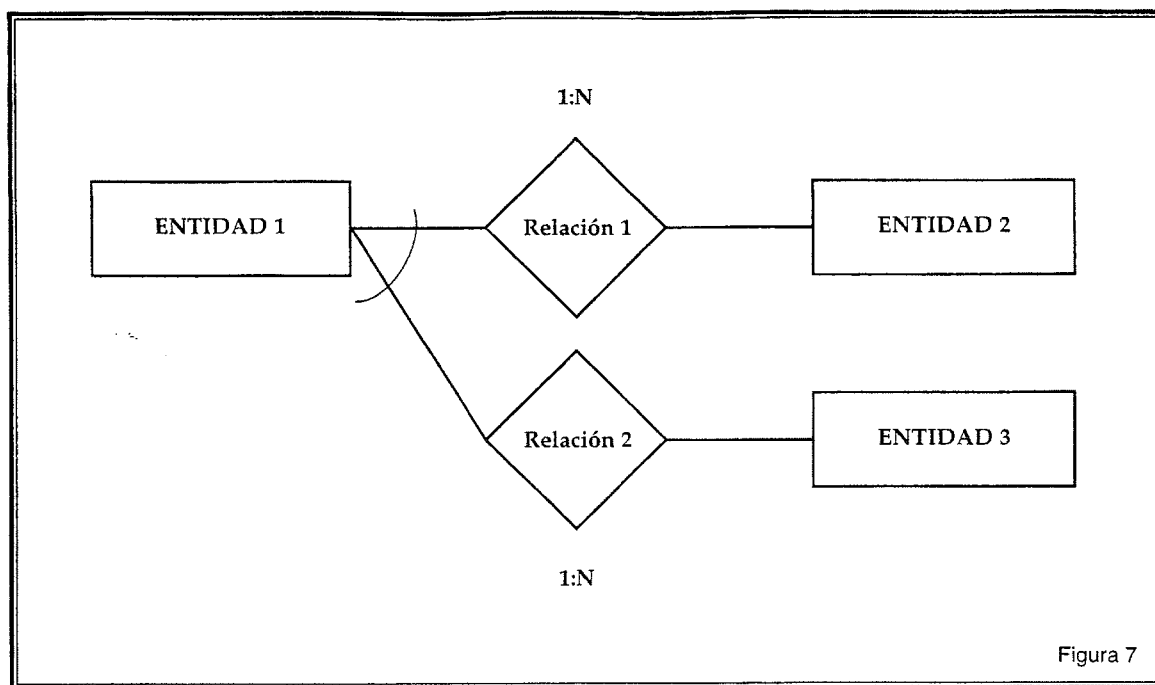


Figura 7

La Entidad 1, que participa en dos asociaciones (Relación 1 y Relación 2), lo hace de forma exclusiva si una instancia de la Entidad 1, participa en una y sólo una de la asociaciones, ya sea la Relación 1 o la Relación 2.

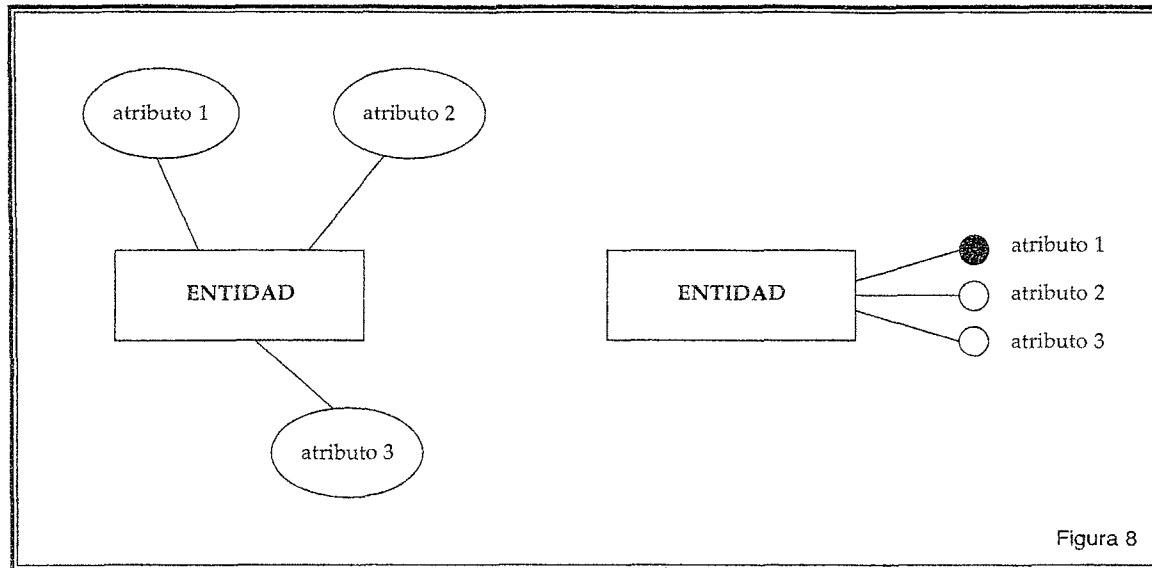
3.1.3. Atributo.

Se define atributo como cada una de las propiedades, características o unidades de información básicas de una entidad o de una asociación o interrelación.

Se entiende por identificador principal o clave aquel o aquellos atributos de una entidad que identifican unívocamente cada una de las ocurrencias de la misma. Por ejemplo, dada la entidad «Empleado» y los atributos DNI, Nombre, Apellidos, Domicilio y Edad, el atributo DNI es el identificador principal.

Se define dominio como el conjunto de valores que puede tomar un atributo dentro del universo del discurso que se está modelizando.

Aunque no es muy corriente, los atributos se pueden representar gráficamente, si bien sólo se representará el identificador principal o clave. La forma de hacerlo es mediante un círculo etiquetado con el nombre de la clave, que deberá ser único e inequívoco. El círculo se unirá a la entidad o asociación que corresponda mediante una línea continua.



3.1.4. Restricciones o limitaciones de integridad.

Las restricciones a considerar en el modelo E/R, también conocidas como limitaciones de integridad, pueden afectar a los atributos, a las asociaciones o interrelaciones y a las entidades. Distinguiremos dos tipos de restricciones:

1. Restricciones sintácticas. Afectan a los atributos y hacen referencia al formato de escritura del atributo. Por ejemplo, si se trata del atributo «fecha», podremos decir que en el modelo que estamos representando su formato debe ser AAAA/MM/DD.
2. Restricciones semánticas. Es una proposición que puede tomar el valor de «cierto» o «falso» para cada ocurrencia de una entidad. Dicho de otra forma, las restricciones semánticas definen los valores de los atributos permitidos o prohibidos para cada instancia de una entidad-tipo o asociación-tipo. Se clasifican en:
 - Restricciones estructurales, que son inherentes al universo del discurso objeto de modelización, es decir, al modelo en sí. Por ejemplo, el atributo «mes» de la entidad «Vacaciones» sólo podrá tomar uno de los doce valores posibles de los meses del año.
 - Restricciones de usuario, que son específicas de la modelización concreta que se realiza del universo del discurso. Este tipo de restricciones pueden definirse sobre:
 - Valores, delimitando los valores posibles de un atributo perteneciente a un elemento del modelo. Por ejemplo, el atributo «salario» de la entidad «Empleado» no puede tomar un valor superior a 5.000.000.
 - Asociaciones, delimitando, o bien el grado de una asociación, esto es, el número de entidades que pueden intervenir en una asociación o interrelación, o bien, la cardinalidad máxima o tipo de correspondencia de la asociación, o bien la cardinalidad individual de cada entidad participante en la asociación.

- **Cardinalidad funcional**, determinando la existencia de una dependencia funcional entre entidades de una asociación. Por ejemplo, en la asociación «dirige» entre las entidades «Empleado» y «Departamento», si se establece que la cardinalidad de la entidad «Empleado» es (0,1), se puede establecer la restricción de que existe una dependencia funcional de «Empleado» sobre «Departamento» a través de la asociación «dirige», que además es débil ya que la cardinalidad mínima de «Empleado» es cero.

En definitiva, las restricciones o limitaciones de integridad aportan más semántica al modelo conceptual, asegurando la coherencia y cohesión de los datos que se recogen e informando sobre los estados y transiciones prohibidos en el universo del discurso.

3.2. ETAPAS DEL MODELADO CONCEPTUAL DE DATOS.

Puesto que el modelado conceptual de datos pretende recoger y representar toda la información del mundo real que es relevante para el sistema, sin considerar cómo se va a implementar el mismo, es normal admitir, tal como sucede con el modelo E/R, que se puedan representar todo tipo de grados de las asociaciones o relaciones y todo tipo de cardinalidades. Sin embargo, hay que matizar que no existe un algoritmo que permita de forma normalizada elaborar el modelo conceptual de datos a través del formalismo del modelo E/R.

No obstante, y dada la importancia fundamental que tiene el modelado conceptual de datos para el desarrollo de un sistema de información, se indicará una serie de pautas y recomendaciones que sirvan de ayuda para su realización. En este sentido, las etapas en que puede subdividirse el modelado conceptual de datos son las siguientes:

1. Análisis de requisitos.

Esta primera etapa, en general común para datos y procesos, es la etapa de percepción, identificación y descripción de los fenómenos del mundo real a analizar, y ha de responder a la pregunta: «¿Qué representar?».

A estos efectos, se parte del «real percibido» por el usuario, esto es, de la visión que tiene éste del universo del discurso que va a ser objeto del sistema, y mediante el estudio de las reglas de la organización, entrevistas a los usuarios y el análisis de la documentación que se pueda aportar (listados, pantallas, normativas, etc.) se llega a elaborar un esquema descriptivo de la realidad, que se expresa, generalmente, en lenguaje natural.

2. Etapa de conceptualización.

En esta etapa, que responde a la pregunta: «¿Cómo representar?», se transforma el esquema descriptivo, refinándolo y estructurándolo adecuadamente. Será necesario, pues, buscar una representación normalizada que se apoye en un modelo de datos que cumpla determinadas propiedades (coherencia, simplicidad, no redundancia, etc.) para llegar así al denominado esquema conceptual. Un modelo de datos que cumple tales requisitos es el modelo E/R

Para obtener el esquema conceptual será necesario, entonces, interpretar las frases del lenguaje natural transformándolas en los elementos del modelo E/R, es decir, en entidades, relaciones, atributos y restricciones semánticas.

A este respecto, algunas reglas que ayudan a elaborar un primer esquema conceptual partiendo de lo expresado en lenguaje natural, son las siguientes:

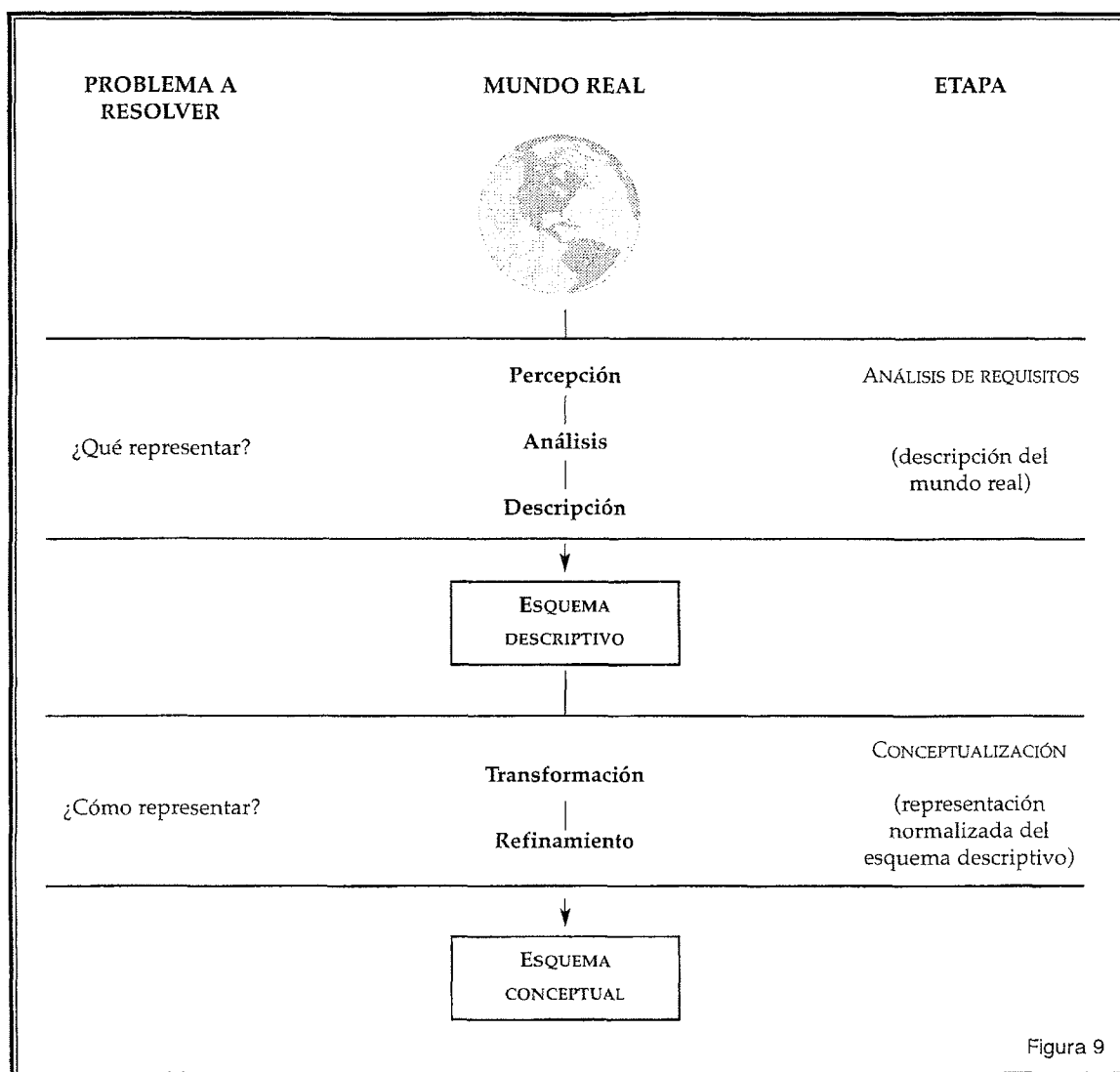
- Un nombre común que actúa como sujeto o complemento directo en una frase, generalmente es una entidad, o en su defecto, un atributo. Por su parte, los nombres propios suelen indicar instancias u ocurrencias de una entidad. Por ejemplo: «Los empleados trabajan en proyectos, en concreto, Juan trabaja en dos proyectos». Empleado y Proyecto son entidades y Juan es una ocurrencia de Empleado.
- Un verbo es una asociación o relación. En el ejemplo anterior, trabajar indica una asociación entre las entidades Empleado y Proyecto.
- Una preposición o una frase preposicional entre dos nombres suele indicar una asociación o relación, o bien, la asociación entre una entidad y sus atributos. Por ejemplo, en la frase «los hijos del empleado...» se está estableciendo una asociación entre las entidades Empleado e Hijo, mientras que en la frase «el domicilio del empleado...», se está asociando el atributo domicilio a la entidad Empleado.

Este proceso se realiza de forma iterativa hasta que se introducen y clasifican todos los objetos del universo del discurso de forma satisfactoria. Para ello se podrá seguir una de estas cuatro estrategias:

- Descendente, que consiste en refinar los conceptos de forma progresiva. Parte de una única entidad que describe el universo del discurso y se va descomponiendo sucesivamente con mayor nivel de detalle. Su mayor inconveniente es que requiere un alto nivel de abstracción, por contra, no presenta ningún problema secundario.
- Ascendente, esto es, ir construyendo los conceptos como síntesis de otros más elementales. Aquí se parte del nivel más bajo, que son los atributos, y se van agrupando éstos en entidades. Posteriormente se crean las asociaciones entre entidades hasta obtener el esquema completo. Esta estrategia facilita las decisiones de diseño, pero tiene el inconveniente de que es necesario reestructurar el modelo constantemente.
- Horizontal, también llamada «inside-out» o método de la mancha de aceite. En esta estrategia se parte de un modelo E/R muy elemental y se va completando a medida que se deducen más cosas del esquema escrito en lenguaje natural. Su gran ventaja es que facilita la descripción de nuevos conceptos relacionados con los anteriores, pero tiene el inconveniente de que sólo se obtiene una visión global del modelo al final.
- Mixta. Consiste en dividir el problema en modo descendente, ascendente y horizontal, con lo que aprovecha las ventajas de cada estrategia. Su único inconveniente es que requiere decisiones críticas acerca del esqueleto inicial del esquema.

Una vez obtenido el esquema conceptual (modelo conceptual de datos) mediante la técnica del modelo E/R, se deberá analizar la existencia de redundancias (atributos redundantes y asociaciones redundantes) y eliminarlas, ya que pueden crear problemas en la implementación de la base de datos. Dicho de otra forma, una vez obtenido el esquema conceptual de datos habrá que convertir éste en un esquema lógico de datos (modelo lógico), si bien, como ya se ha indicado, esto corresponde al proceso de Diseño del Sistema de Información.

En la siguiente figura se representa el proceso de modelización conceptual de datos.



3.3. EJEMPLO DE APLICACIÓN DE LA TÉCNICA DEL MODELO E/R.

Al objeto de reafirmar los conocimientos sobre los elementos que componen el modelo E/R de Peter Chen, vamos a representar gráficamente dicho modelo a través del siguiente ejemplo.

BIBLOS es una Asociación de Bibliotecas integrada por veinte centros. Cada centro (biblioteca), además de llevar la constatación de los libros que posee y de los autores de éstos, concede préstamos a sus usuarios con la condición de que cada uno de ellos sólo puede tener concedido en cada momento un préstamo vivo, si bien, éste puede constar de hasta tres libros.

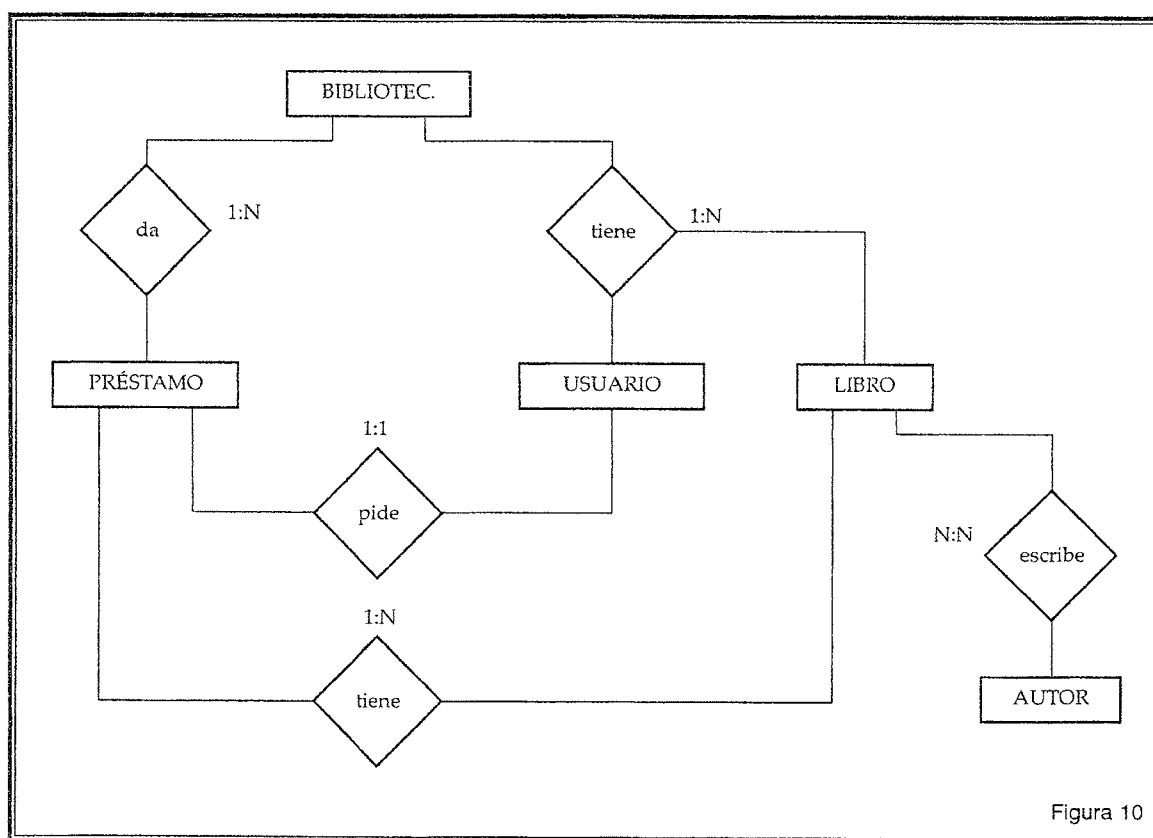
Se pide representar el modelo conceptual de datos en la forma de modelo E/R de Chen para la organización BIBLOS.

Solución.

Lo primero es identificar las entidades en nuestro universo del discurso, es decir, todo aquello que es relevante para el sistema y sobre lo que se desea guardar información. Las entidades identificadas son las siguientes:

- Biblioteca. Cada uno de los veinte centros que integran la Asociación «BIBLOS». La clave o identificador principal de cada instancia u ocurrencia de la entidad será el código de la biblioteca. Llamémosle Cod-Bibl.
- Libro. Cada uno de los ejemplares que tiene una biblioteca, aunque tengan el mismo título. La clave de cada instancia será Cod-Lib.
- Usuario. Cada socio de una biblioteca. La clave de cada instancia será el NIF.
- Préstamo. Los préstamos vivos (no cancelados) que realiza una biblioteca a cada uno de sus usuarios. La clave de cada instancia será el número de préstamo Núm-Pres.

De las condiciones dadas, la representación gráfica del modelo conceptual de datos, siguiendo las reglas del modelo E/R de Chen, y su explicación es la que figura a continuación.



Cada «biblioteca» puede conceder varios préstamos, pero cada «préstamo» sólo puede ser concedido por una biblioteca (relación 1:N). De la misma forma, cada «biblioteca» puede tener muchos usuarios y muchos libros (la interrelación «tiene» es ternaria puesto que participan de ella tres entidades), ahora bien, un «usuario» o un «libro» pertenecen sólo a una biblioteca (relación 1:N en ambos casos).

Por otra parte, un «usuario» puede tener concedido un «préstamo vivo», pero éste sólo se concede a un usuario (relación 1:1). Asimismo, un «prestamo» puede contener varios libros, pero un «libro» sólo puede estar contenido en un préstamo (relación 1:N). Finalmente, un «libro» puede estar escrito por varios autores, y un «autor» puede haber escrito varios libros (relación N:N).

En cuanto a la cardinalidad mínima y máxima de las entidades, de las condiciones del enunciado se deduce lo siguiente:

- Asociación «Biblioteca» tiene «Usuarios» y «Libros». Una biblioteca tiene como mínimo un usuario y un libro y como máximo N. Asimismo, un usuario o un libro sólo pertenecen a una biblioteca. Por tanto, las cardinalidades serán BIBLIOTECA (1, 1), USUARIO (1, N) y LIBRO (1, N).
- Asociación «Biblioteca» concede «Préstamo». Una biblioteca puede no tener concedido ningún préstamo, o tener concedidos N. Por otra parte, un préstamo sólo es concedido por una biblioteca. Por tanto, BIBLIOTECA (1, 1) y PRÉSTAMO (0, N).
- Asociación «Usuario» tiene concedido «Préstamo». Un usuario puede no tener concedido ningún préstamo, o tener concedido 1. Por otra parte, un préstamo sólo puede ser concedido a un usuario. Por tanto, USUARIO (1, 1) y PRÉSTAMO (0, 1).
- Asociación «Préstamo» contiene «Libro». Un préstamo contiene como mínimo un libro y como máximo tres. Por otra parte, un libro puede no pertenecer a ningún préstamo o como máximo a uno. Por tanto, LIBRO (1, N) siendo $N = 3$ y PRÉSTAMO (0, 1).

(Tanto en esta asociación como en las dos anteriores, la entidad «Préstamo» es una entidad débil ya que su cardinalidad mínima es cero, lo que significa que su existencia no se debe a sí misma sino a su relación con las otras entidades. Estas asociaciones, por relacionar una entidad regular con una débil, tienen la consideración de asociación débil).

- Asociación «Libro» escrito por «Autor». Un libro puede haber sido escrito por un autor, como mínimo, o por N. Asimismo, un autor puede haber escrito un libro como mínimo, o N. Por tanto, LIBRO (1, N) y AUTOR (1, N).

Por no complicar el dibujo del modelo, no se ha representado la clave o identificador principal de las entidades, ni la cardinalidad de las mismas.

¡ADVERTENCIA!: EL QUE EL MODELO CONCEPTUAL SEA ÚNICO PARA UN DETERMINADO SISTEMA DE INFORMACIÓN NO QUIERE DECIR QUE SU GRAFISMO NO PUEDA SER DIFERENTE. EVIDENTEMENTE, CADA PERSONA LO INTERPRETARÁ SEGÚN SU PUNTO DE VISTA, Y ASÍ LO EXPRESARÁ. AHORA BIEN, CON INDEPENDENCIA DE CÓMO SE REPRESENTA, LO IMPORTANTE ES QUE ES INDEPENDIENTE DE LAS CARACTERÍSTICAS DEL ENTORNO FÍSICO (EQUIPO) EN QUE SE VAYA A IMPLANTAR EL SISTEMA.

4. EL MODELO ENTIDAD/RELACIÓN EXTENDIDO.

Desde su presentación en 1976, el modelo E/R de Chen ha recibido aportaciones de diversos autores con el fin de dotarlo de una mayor capacidad semántica, que permita reflejar más fielmente el mundo real, dando lugar a lo que se ha venido en llamar Modelo Entidad/Relación Extendido (EE/R).

Las principales aportaciones del Modelo EE/R respecto al de Chen son las siguientes:

1. El modelo EE/R, respecto al tipo de correspondencia de una relación, considera no sólo la cardinalidad máxima, sino también la cardinalidad mínima, que se define como el número mínimo de ocurrencias de una entidad-tipo que se pueden interrelacionar con cada ocurrencia de las otras entidades-tipo en la asociación que se está considerando.

2. En relación con las dependencias entre entidades, el modelo EE/R considera dos tipos fundamentales de dependencias:

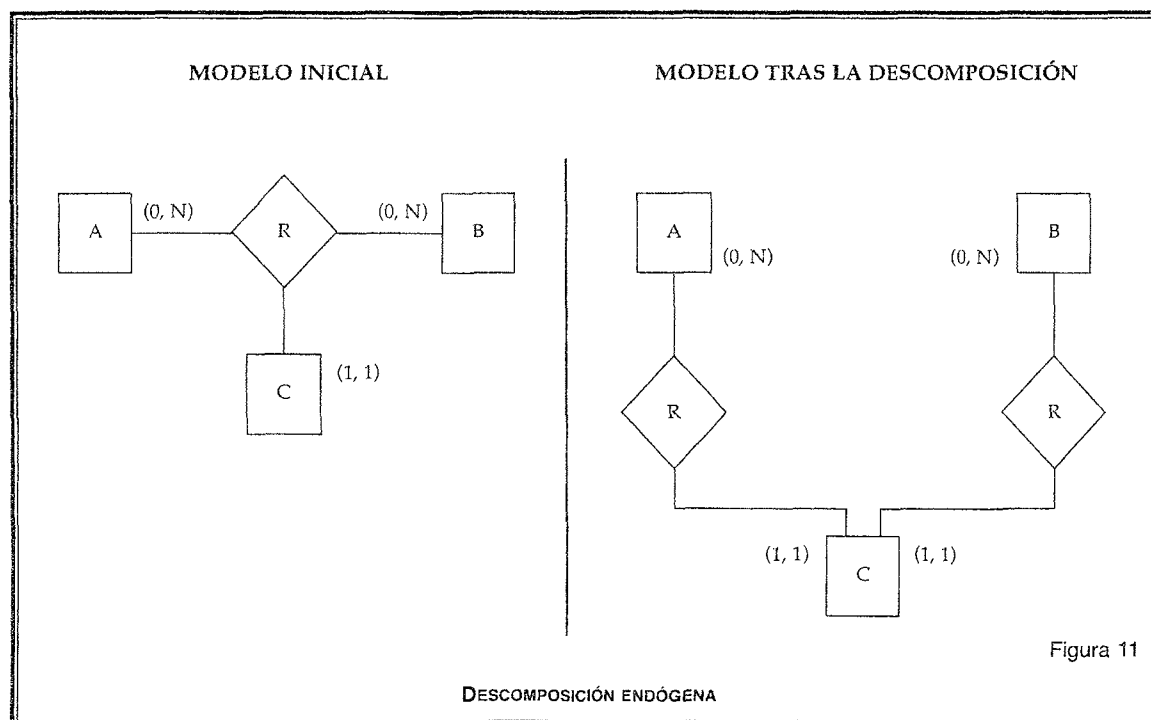
- Dependencia en existencia: una entidad-tipo depende en existencia de otra cuando las ocurrencias de la primera no tienen sentido sin las ocurrencias de la segunda. Esta dependencia configura la entidad-tipo como entidad débil, frente a las otras entidades-tipo que se caracterizan como regulares.
- Dependencia en identificación: en el caso de una entidad-tipo débil, se dice que además tiene dependencia en identificación si no se puede identificar sólo con sus propios atributos, sino también a partir de los atributos de las asociaciones en las que participa.

3. El modelo EE/R considera también el concepto de atributo derivado, definiéndolo como aquél que puede ser deducido de forma automática a partir de otros de la entidad-tipo o asociación-tipo correspondiente, siendo por lo tanto redundante su almacenamiento. Se representan en el modelo mediante una etiqueta asociada a la línea del atributo.

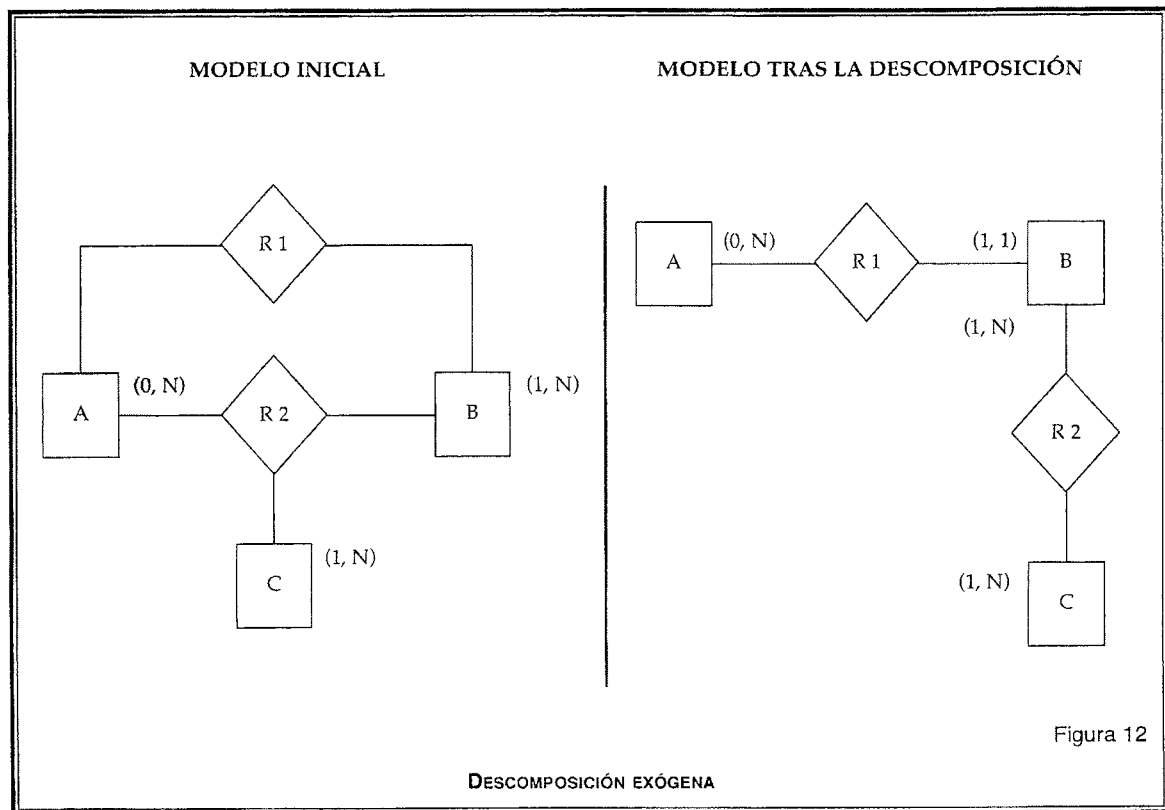
Será en el momento del diseño técnico de la base de datos del sistema de información cuando habrá que decidir si se almacena el atributo por razones de eficiencia.

4. Otro concepto que introduce el modelo EE/R es el de descomposición de las asociaciones. Su objeto es simplificar el modelo conceptual, manteniendo su semántica, para hacerlo más legible y fácil de comprender. Existen dos tipos de descomposición de las asociaciones:

- Descomposición endógena. «Si en una asociación o interrelación de grado mayor que dos, una entidad tiene una cardinalidad individual (1,1), la asociación se puede descomponer en tantas asociaciones binarias (grado 2) como entidades participen menos uno, y los atributos de la asociación o interrelación original pasan a la entidad con cardinalidad (1,1)».



- Descomposición exógena. «Cuando existen varias entidades de una asociación que participan en otra en la cual una de esas entidades determinan a otra, es decir, existe una dependencia funcional entre ellas, si los subconjuntos de entidades involucrados en ambas asociaciones son los mismos, la entidad determinada en la segunda asociación puede desaparecer de la primera».

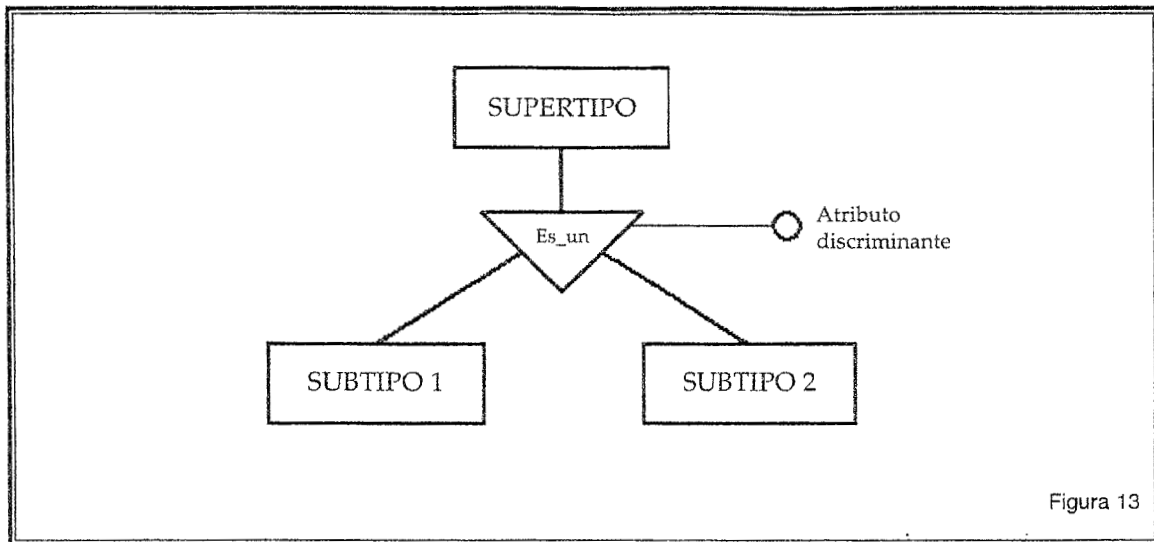


5. Por último, el modelo EE/R permite definir el concepto de subentidad, como un conjunto de ocurrencias de una entidad y establecer así relaciones jerárquicas.

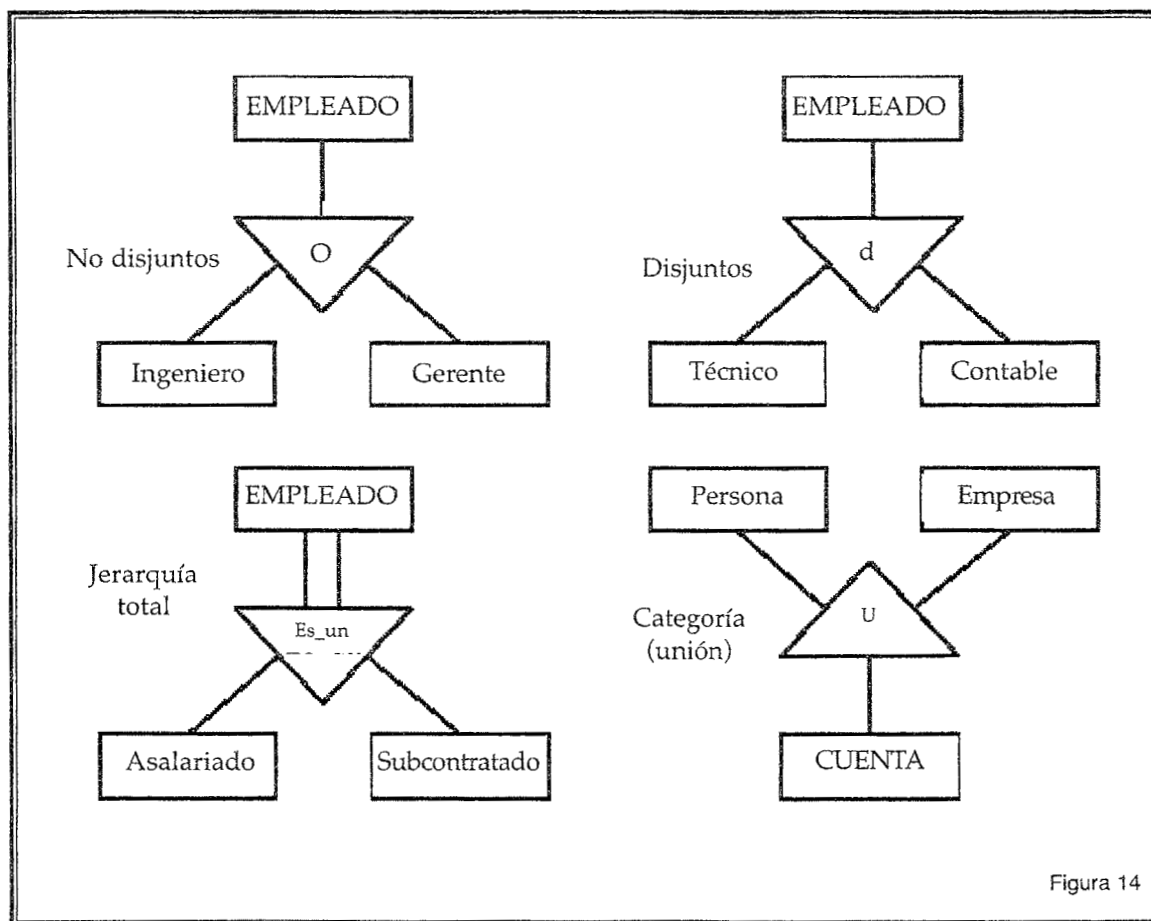
El concepto de subentidad o subtipo permite definir la idea de herencia y aproxima el modelo EE/R al paradigma de la Orientación a Objetos. El concepto de herencia se plasma en que toda ocurrencia de una subentidad es ocurrencia de la entidad asociada o supertipo, y todo atributo de ésta es un atributo de la subentidad. Sin embargo, ambas características no se dan en sentido inverso.

La representación de las jerarquías se realiza mediante un triángulo invertido (etiquetado como «es un»), con la base paralela al rectángulo que representa el supertipo y conectando a éste y a los subtipos. Si la división en subtipos viene determinada en función de los valores de un atributo discriminante, éste se representará asociado al triángulo que representa la relación.

Otra forma de representación de esta jerarquía es la inclusión de los rectángulos que representan a las subentidades dentro del que representa a la entidad.



En el triángulo se representará: con una letra d el hecho de que los subtipos sean disjuntos, con un círculo o una O si los subtipos pueden solaparse y con una U el caso de uniones por categorías. La presencia de una jerarquía total se representa con una doble línea entre el supertipo y el triángulo.



La existencia de supertipos y subtipos, en uno o varios niveles, da lugar a una jerarquía, que permitirá representar una restricción del mundo real. Las relaciones jerárquicas más importantes son:

- La **generalización**, que permite abstraer un tipo de entidad de nivel superior (supertipo) a partir de varios tipos de entidad (subtipos); en estos casos los atributos comunes y relaciones de los subtipos se asignan al supertipo. Se pueden generalizar por ejemplo los tipos profesor y estudiante obteniendo el supertipo persona.
- La **especialización**, que es la operación inversa a la generalización, en ella un supertipo se descompone en uno o varios subtipos, los cuales heredan todos los atributos y relaciones del supertipo, además de tener los suyos propios. Un ejemplo es el caso del tipo empleado, del que se pueden obtener los subtipos secretaria, técnico e ingeniero.

Se denomina categoría al subtipo que aparece como resultado de la unión de varios tipos de entidad. En este caso, hay varios supertipos y un sólo subtipo. Si por ejemplo se tienen los tipos persona y compañía y es necesario establecer una relación con vehículo, se puede crear propietario como un subtipo unión de los dos primeros.

- La **agregación**, que consiste en construir un nuevo tipo de entidad como composición de otros y su tipo de relación y así poder manejarlo en un nivel de abstracción mayor. Por ejemplo, se tienen los tipos de entidad empresa y solicitante de empleo relacionados mediante el tipo de relación entrevista; pero es necesario que cada entrevista se corresponda con una determinada oferta de empleo. Como no se permite la relación entre tipos de relación, se puede crear un tipo de entidad compuesto por empresa, entrevista y solicitante de empleo y relacionarla con el tipo de entidad oferta de empleo. El proceso inverso se denomina desagregación.

4.1. UN EJEMPLO DE MODELO ENTIDAD/RELACIÓN EXTENDIDO.

Se presenta a continuación el Modelo Entidad-Relación Extendido para un sistema de gestión de técnicos y su asignación a proyectos dentro de una empresa u organización.

Como se aprecia en el diagrama, TÉCNICO es un subtipo de EMPLEADO, generado por especialización, pues era necesario para establecer la relación Trabaja en con PROYECTO, ya que no todos los empleados de la empresa, como los administrativos, son susceptibles de trabajar en un proyecto. La entidad TÉCNICO tendrá los atributos de EMPLEADO más el atributo nivel.

Los tipos de correspondencia son 1:N entre DEPARTAMENTO y EMPLEADO, pues un departamento tiene 1 o varios empleados. Entre TÉCNICO y PROYECTO es M:N, pues un técnico puede trabajar en 1 o varios proyectos, y en un proyecto trabajan 1 o varios técnicos.

Por otro lado, se han incluido atributos que caracterizan la relación Trabaja en, como son fecha de asignación y fecha de cese, ya que un técnico no siempre estará trabajando en un proyecto, sino en determinado período.

(Nota. Esta notación es la más habitual, pero Métrica Versión 3 no exige su utilización).

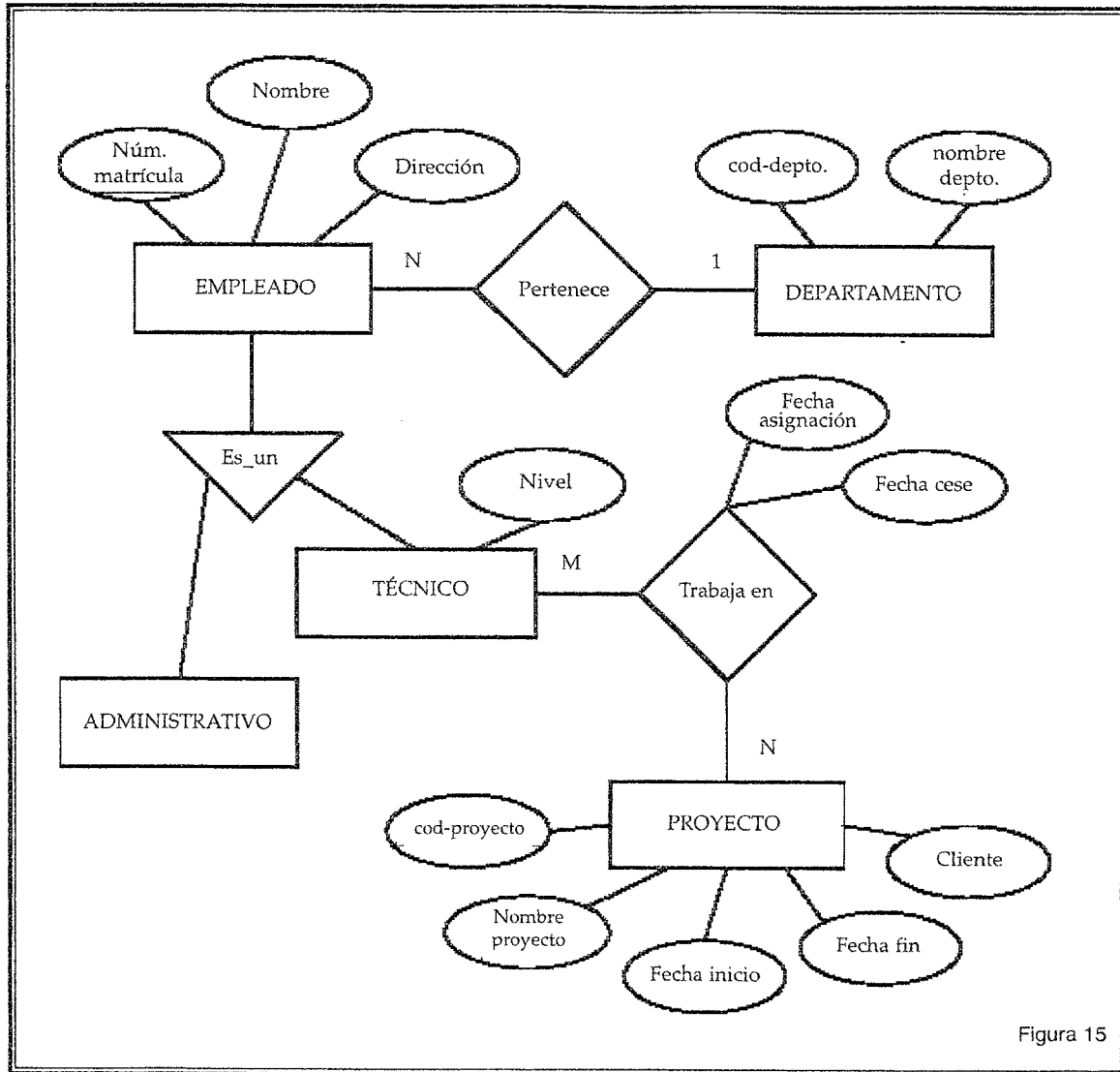


Figura 15

BIBLIOGRAFÍA

- Ingeniería del Software. Un enfoque práctico. Roger S. Pressman. Ed. McGraw-Hill.
- Concepción y Diseño de Bases de Datos. Adoración de Miguel y Mario Piattini. Ed. Ra-Ma.
- Análisis y Diseño detallado de Aplicaciones Informáticas de Gestión. Mario Piattini y otros. Ed. Ra-Ma.
- Ingeniería del Software. Ian Sommerville. Ed. Addison-Wesley.
- Metodología de Planificación, Desarrollo y Mantenimiento de Sistemas de Información. Métrica versión 3. Técnicas y Prácticas. Ministerio para las Administraciones Públicas.
- Metodología de Planificación y Desarrollo de Sistemas de Información. Métrica versión 2.1. Guía de Técnicas. Ministerio para las Administraciones Públicas. Ed. Tecnos.

- Temario de las pruebas selectivas para ingreso en el Cuerpo Superior de Sistemas y Tecnologías de la Información de la Administración del Estado. ASTIC.
- Temario de las pruebas selectivas para el acceso, por promoción interna, al Cuerpo de Gestión de Sistemas e Informática de la Administración del Estado. Ministerio para las Administraciones Públicas.
- Temario del Máster en Ingeniería del Software. Facultad de Informática. Universidad Politécnica de Madrid. Ed. Estudios Financieros.



11

12

13

14