

Indice

MODELIZACION CONCEPTUAL DE DATOS

MODELOS DE DATOS: CONCEPTOS GENERALES
 REPRESENTACIÓN SIMBÓLICA DE LA INFORMACIÓN
 MODELOS DE DATOS
 POTENCIA SEMÁNTICA
 FILOSOFÍA DE LA MODELIZACIÓN CONCEPTUAL

MODELO ENTIDAD-INTERRELACION

ENTIDADES Y TIPO DE ENTIDADES
 INTERRELACIONES Y TIPOS DE INTERRELACIONES
 DIAGRAMAS E-R
 RESTRICCIONES DE CARDINALIDAD DE CORRESPONDENCIA
 RESTRICCIONES DE DEPENDENCIA EXISTENCIAL
 CLAVES PRIMARIAS - RESTRICCIONES DE IDENTIFICACION
 GENERALIZACION Y ESPECIALIZACION
 AGREGACION
 CASO DE ESTUDIO
 Definición de tipos de entidades
 Definición de tipos de interrelaciones
 Identificación de atributos para tipos de entidades
 Identificación de atributos para tipos de interrelaciones
 Identificación de claves primarias para tipos de entidades
 Identificación de claves primarias para tipos de interrelaciones
 Diagrama ENTIDAD-INTERRELACION

MODELO DE DATOS RELACIONAL

INTRODUCCION
 CONCEPTOS BASICOS DEL MODELO
 DOMINIO
 RELACION
 ATRIBUTO
 ESQUEMA DE RELACION
 SUPERCLAVE, CLAVE CANDIDATA, CLAVE PRIMARIA, CLAVE FORANEA
 SUPERCLAVE
 CLAVE CANDIDATA: UNICIDAD - MINIMALIDAD
 CLAVE PRIMARIA
 CLAVE FORANEA
 MODELO RELACIONAL: RESTRICCIONES
 REGLA DE INTEGRIDAD DE ENTIDAD
 REGLA DE INTEGRIDAD REFERENCIAL
 ALGEBRA RELACIONAL
 SELECCIÓN o RESTRICCION
 PROYECCION
 UNION
 DIFERENCIA
 PRODUCTO CARTESIANO
 OPERACIONES ADICIONALES
 JOIN
 INTERSECCION
 DIVISION
 COMPOSICION DE RELACIONES

APENDICE

TRANSFORMACION DEL MODELO E-R A OTROS MODELOS IMPLEMENTABLES
 TRANSFORMACION DE UN MODELO E-R A TABLAS
 TIPOS DE ENTIDADES REGULARES
 TIPOS DE ENTIDADES DÉBILES
 TIPOS DE INTERRELACIONES
 GENERALIZACION Y ESPECIALIZACION

MODELIZACION CONCEPTUAL DE DATOS

MODELOS DE DATOS: Conceptos Generales

Para que los datos sean útiles proveyendo información, deben ser organizados de manera tal que puedan ser procesados efectivamente.

En la modelización de datos tratamos de organizar los datos de manera que representen lo más cercanamente posible la situación del mundo real, siendo dicha representación manejable por computadoras; en general, estos dos requerimientos resultan conflictivos.

Dado que no es posible tener un conocimiento completo del mundo real, debemos centrar nuestra atención en el significado de los datos apropiado para su utilización, ocultando o desconociendo otros significados que resulten irrelevantes o inadecuados para el propósito requerido.

Se trata, entonces, de comprender las características de los datos que son importantes para capturar la esencia de su significado. Estas características permiten establecer cómo van a ser organizados y procesados, y pueden ser descriptas a través de enunciaciones generales. Un conjunto formal y consistente de tales enunciaciones define un modelo de datos.

En términos generales, podemos decir que un **MODELO DE DATOS** es una *herramienta intelectual que provee una interpretación del mundo real*, lo suficientemente poderosa para ayudar a entender cómo están relacionados los datos, y lo suficientemente *abstracta*, para ser mínimamente perturbable ante los cambios concernientes al aspecto evolutivo del mundo real.

Es un *dispositivo de abstracción*, que nos permite ver el contenido de información de los datos, en lugar de sus valores individuales.

Los MODELOS son usados extensivamente en muchas disciplinas para mejorar la comprensión y abstraer los detalles, entre otros podemos mencionar MODELOS FÍSICOS, MATEMÁTICOS, ECONÓMICOS, etc. Nuestro tema de interés se limita exclusivamente a los MODELOS DE DATOS, los cuales pueden ser codificados y manipulados fácilmente por computadoras.

Si intentamos determinar, al menos aproximadamente, qué es un dato, podemos suponer como una definición de trabajo que una pieza atómica de dato es la siguiente tupla:

< nombre de objeto, propiedad del objeto, valor de la propiedad del objeto, instante >

dado que un fenómeno o idea, usualmente refiere a *un objeto* y a algún *aspecto del objeto*, el cual captura un determinado *valor* en un cierto *momento*.

Sin embargo, el tiempo real implica cierta sincronización entre los fenómenos que no es del todo representable, en cambio nos interesa a menudo el tiempo relativo del fenómeno, lo cual puede ser adecuadamente capturado, ordenando los fenómenos antes que registrando el tiempo real. Podríamos entonces dejar de lado la noción de tiempo y reemplazarla por otras propiedades explícitas que permitan establecer el ordenamiento entre objetos.

Entonces nuestra definición de *dato elemental* se convertiría en la terna:

< nombre de objeto, propiedad del objeto, valor de la propiedad del objeto >

Hay muchas maneras en que estos aspectos de los datos pueden ser representados y relacionados, dando lugar a distintos MODELOS DE DATOS.

Una manera simple y potente, consiste en categorizar los datos, de modo tal que datos puestos en la misma categoría manifiesten similitudes; en consecuencia podemos establecer que esas similitudes constituyen propiedades de la categoría y, por lo tanto, son heredables por los datos que pertenecen a la categoría. Hacemos una clara distinción entre *datos* y *categorías de datos*.

Una desventaja de esta clase de MODELOS DE DATOS, que damos en llamar *tipificados estrictamente* (*strictly typed*), es que obligan a “encajar el mundo real” de modo que cada dato caiga dentro de una categoría; esta falta de flexibilidad, algunas veces, presenta dificultades para representar diferencias semánticas sutiles, o, en el peor de los casos, resultan datos del mundo real que no caben en categoría alguna y, por lo tanto, no pueden ser manejados en el MODELO DE DATOS.

Por otra parte, podemos justificar esta imposición, teniendo en cuenta la cantidad de objetos a tratar y las limitaciones de la inteligencia humana. Dado que en la modelización de datos se debe tratar con muchos objetos, es necesario cierto marco restrictivo que los comprenda en un conjunto limitado de ideas que resulten más fáciles para el entendimiento.

Concluyendo, la gran ventaja, entonces, es que las propiedades de los datos pueden ser abstraídas e investigadas en términos de sus categorías. Es decir, una teoría puede ser formulada en base a categorías que comprendan las propiedades de los datos en sí; luego, probar algo acerca de las categorías implica haberlo probado acerca de los datos.

REPRESENTACIÓN SIMBÓLICA DE LA INFORMACIÓN

Conocimiento es el entendimiento de la naturaleza, cualidades y relaciones de las cosas. Cuando percibimos u observamos un fenómeno, obtenemos una porción incremental de conocimiento. Denominamos *información* a estos incrementos de conocimiento.

La información valiosa debe ser registrada en una forma simbólica inteligible para que pueda ser comunicada a otras personas. Para estos fines resulta de importancia fundamental contar con una adecuada representación simbólica de la información.

La pieza más elemental de información es el *dato* (ej., una cifra de sueldo). Un dato no es de utilidad por sí mismo, si no está vinculado con otros datos (ej., la identificación de un empleado), por lo que el almacenamiento y la comunicación de información requieren la estructuración de datos. La estructura y el contexto le dan *significado* a los datos posibilitando su entendimiento.

El lenguaje natural es nuestra forma primaria de representación y comunicación de información. Si bien es el medio más generalizado para ese propósito, no siempre constituye el mejor medio. En muchas situaciones resulta más útil establecer formas especializadas de representación de la información. Las fórmulas matemáticas, los mapas carreteros y las partituras musicales, constituyen ejemplos de *modelos* especializados de representación simbólica de la información.

Los datos de una base de datos se almacenan y procesan en un computador por medio de Sistemas de Gestión de Bases de Datos (SGBD o DBMS -Data Base Management Systems-). Se han desarrollado diferentes Modelos de Datos (Relacional, Jerárquico, Red) para la representación de la información en los sistemas computarizados. En este contexto, un Modelo de Datos describe la forma de representación de la información utilizada por un SGBD en particular. Así, decimos que un SGBD Relacional (RDBMS) responde a un Modelo de Datos Relacional.

Estos Modelos de Datos, que describen cómo se almacenan los datos en el computador, resultan inadecuados para representar el *significado* de los datos almacenados. Por este motivo han surgido metodologías de diseño de bases de datos sustentadas en la utilización de Modelos de Datos de mayor potencia semántica orientados a la descripción de la realidad que estará representada en las bases de datos. Estos Modelos, que se denominan *modelos conceptuales* o *semánticos*, tienden a describir la realidad captando en mayor medida el significado de los datos y de sus vinculaciones en el mundo real.

MODELOS DE DATOS

Podemos decir que un MODELO DE DATOS es una herramienta formal para representar la realidad. Los Modelos de Datos en general, especifican:

- la **estructura** de los datos.
- **restricciones** sobre las estructuras.
- **operaciones** sobre las estructuras.

Un Modelo de Datos establece reglas para la estructuración de los datos (ej, tablas, árboles, redes, etc.), permite establecer restricciones sobre las estructuras para que los datos mantengan su coherencia informativa o *integridad semántica* y define la naturaleza de las operaciones permitidas sobre los datos, dentro del contexto provisto por las estructuras.

Las restricciones pueden abarcar una gran variedad de aspectos, contribuyendo a la descripción más ajustada de la realidad, a través de la exclusión de lo imposible. Por ejemplo, pueden especificarse restricciones provenientes del mundo real tales como:

- los empleados no pueden ganar más que sus jefes.
- para inscribirse en un curso, un estudiante debe satisfacer ciertos requisitos.
- la cantidad de pacientes internados en un hospital no puede exceder la cantidad de camas disponibles.

Por otra parte, además de esta clase de *restricciones* que llamamos *explícitas*, un Modelo de Datos introduce ciertas *restricciones inherentes* como consecuencia de sus propias características y limitaciones de modelización.

POTENCIA SEMÁNTICA

La *potencia semántica* de un Modelo de Datos está determinada por su capacidad de representar, además de la estructura de los datos, *el significado de los mismos y de sus interrelaciones*. Por ejemplo, en la frase “Su altura es de 5 metros”, el dato es el número “5” y el significado es “altura en metros”.

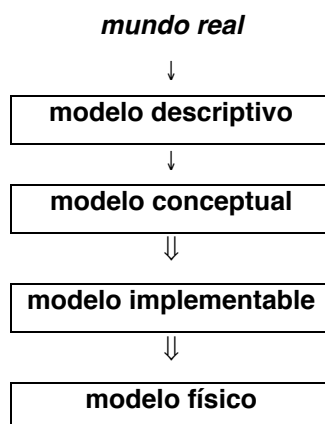
En general, la utilización del computador para el procesamiento de información ha resultado en una separación entre los datos y su significado. Mientras que las computadoras sólo procesan los datos, queda para el usuario o los programas, la tarea de interpretación. A medida que la utilización de computadoras evoluciona hacia aplicaciones más sofisticadas, resulta cada vez más necesario captar y utilizar activamente el significado de los datos (y de sus vinculaciones).

Consecuentemente, es necesario advertir que al traducir el *modelo conceptual* al modelo de datos del SGBD que se utilizará, inevitablemente habrá pérdida de información semántica.

FILOSOFÍA DE LA MODELIZACIÓN CONCEPTUAL

La siguiente figura pretende reflejar el proceso mental mediante el cual pasamos de una realidad informacional a una representación asimilable por un computador. Es claro que deben existir varios niveles de abstracción entre el usuario final que trata con *objetos* como: vuelos, pilotos, aviones o pasajeros y el computador, que almacena e interpreta bits.

El mundo real	compañía de fabricación; líneas de transportes; entidades bancarias; servicios de telecomunicaciones; ciencias; arte; deportes ...
Modelos conceptuales	Modelos infológicos Modelos semánticos Modelo entidad-interrelación Modelo binario
Modelo de bases de datos	Modelo relacional Modelo CODASYL (en red) Modelo jerárquico Modelos híbridos
Estructuras de datos	árboles equilibrados; archivos secuenciales indexados; pilas; colas; listas invertidas
Almacenamiento físico	bits; bytes; código ASCII; EBCDIC



Algunas misceláneas a considerar en el proceso de modelización conceptual:

- Una base de datos es una colección de hechos registrados que reflejan el estado de ciertos aspectos de interés del mundo real. En todo momento el contenido (*extensión*) de la base de datos representa una instantánea del estado del sistema real. Cada cambio en un valor almacenado en la base de datos refleja un cambio ocurrido o planificado en el sistema real.
- Los Modelos de Datos permiten diseñar la estructura (*intensión o esquema*) de una base de datos, de manera que refleje la estructura del sistema que la misma modeliza.
- Un modelo de datos expresa la estructura y las reglas de gestión de un sistema. Por ejemplo, la estructura organizativa y las reglas de funcionamiento de un hospital.
- Un modelo de datos resulta definido por el sistema que representa.
- Como consecuencia de las consideraciones anteriores, se puede decir que la estructura y las reglas de funcionamiento (de gestión) del sistema real determinan objetivamente la estructura de la base de datos, con independencia del diseñador. El diseñador se limita a decidir el alcance (amplitud) y la granularidad (grado de detalle) del modelo y a modelizar conceptualmente el mundo real. La función del diseñador no es crear una estructura de información, sino descubrir y explicitar aquella que está implícita en la realidad. Del modelo conceptual resultante, se deriva objetivamente la estructura de la base de datos.
- Un ambiente de aplicación (un banco, un hospital, una universidad, una industria, etc.) es un sistema realimentado por información que evoluciona con el tiempo. Es posible identificar en él múltiples flujos (dinero, materiales, personas, equipos, órdenes, etc.) siendo el de información el que interconecta y posibilita el funcionamiento del sistema completo.
- Dentro de un ambiente de aplicación cada grupo, actividad, proceso, persona, función y sistema tiene su propia percepción dinámica de la realidad, su propio vocabulario y su propia valoración de lo que es importante. La empresa típica es una composición de estas visiones.
- El resultado del diseño conceptual es un modelo de datos conceptual construido desde el punto de vista de los usuarios y de la aplicación, con independencia de cómo los datos van a ser almacenados y procesados por el computador. En una etapa posterior de *diseño lógico* se traduce el modelo conceptual al modelo de datos que corresponde al SGBD a utilizar, adaptándolo a dicho SGBD y a los aspectos de procesamiento (procesos, frecuencias, tiempos de respuesta, prioridades, etc.). Finalmente, en la etapa de *diseño físico* se realiza la adaptación al computador contemplando las características y limitaciones del mismo.

MODELO ENTIDAD-INTERRELACION

ENTIDADES Y TIPO DE ENTIDADES

Una *Entidad* es algo que existe (concreto o abstracto) y es distinguible de algo similar. Por ejemplo, Juan Cancela con cédula de identidad 5.789.231 es una entidad pues identifica a una persona en particular del universo. La cuenta número 115 de la sucursal Boedo del Banco de la Nación es una entidad pues identifica unívocamente a una cuenta en particular. Una entidad puede ser algo concreto (una persona, un libro) o abstracto (un concepto).

Un *Conjunto de Entidades* (o *Tipo de Entidad*¹) es un conjunto de entidades del mismo tipo. El conjunto de todas las personas que tienen cuentas en un banco, por ejemplo, puede definirse como el conjunto de entidades CLIENTE o el tipo de entidad CLIENTE. El conjunto de entidades CUENTA o el tipo de entidad CUENTA puede representar el conjunto de todas las cuentas de un banco en particular.

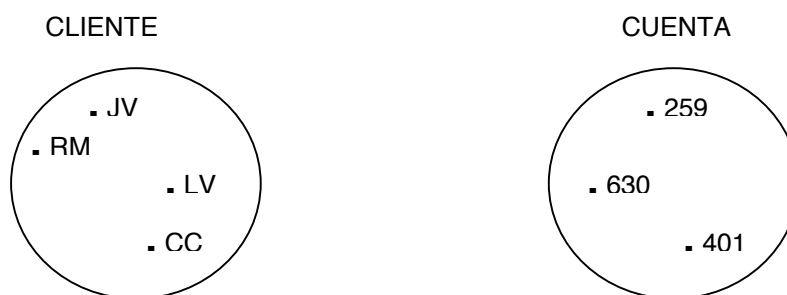


Fig. 1

Existe un predicado asociado a cada conjunto de entidades para probar si una entidad pertenece o no al conjunto. Matemáticamente un conjunto de entidades puede representarse como $E = \{ e / p(e) \}$, donde **E** es el conjunto de entidades **e** que satisfacen el predicado **p**.

Los conjuntos de entidades no necesitan ser disjuntos. Por ejemplo, es posible definir el conjunto de entidades de todos los empleados de un banco (EMPLEADO) y el conjunto de entidades de todos los clientes del banco (CLIENTE). Una entidad PERSONA puede ser una entidad EMPLEADO, una entidad CLIENTE, ambas, o ninguna.

¹ En la literatura se utilizan diferentes nombres para este concepto: conjunto de entidades (entity set), tipo de entidad (entity type), entidad-tipo, entidad genérica. Nosotros usaremos conjunto de entidades o tipo de entidad indistintamente.

Un Tipo de Entidad queda definido por:

- a) un NOMBRE
 - b) un SIGNIFICADO
 - c) ATRIBUTOS
 - d) un IDENTIFICADOR
- a) El NOMBRE debe ser:
- conciso: nombre en singular (Ej.: CLIENTE) o frase simple en singular (CUENTA CORRIENTE)
 - adecuado: de uso común en la organización y que no excluya ciertas ocurrencias
 - único: dentro del modelo evitar sinónimos y homónimos
- b) El SIGNIFICADO o definición debe ser expresado a través de un texto conciso, preciso y claro, indicando, si existieran, dependencias con otros tipos de entidades
- c) Un conjunto de ATRIBUTOS que caracterizan a todas las entidades de ese tipo.
- NOMBRE, CEDULA DE IDENTIDAD, CALLE y CIUDAD son posibles atributos del conjunto de entidades CLIENTE.
 - NUMERO DE CUENTA y SALDO son posibles atributos del conjunto de entidades CUENTA.

Para cada atributo existe un conjunto de valores permitidos, denominado el dominio de ese atributo:

- El dominio del atributo NOMBRE puede ser el conjunto de todos los nombres permitidos de personas.
- El dominio del atributo NUMERO DE CUENTA puede ser el conjunto de todos los números de cuenta permitidos en ese banco.

Formalmente un atributo es una función que hace corresponder un conjunto de entidades a un dominio o producto cartesiano de dominios. Así, toda entidad está descripta por un conjunto de pares (atributo, valor); un par por cada atributo del conjunto de entidades. Por ejemplo, una entidad EMPLEADO en particular, está descripta por el conjunto {(nombre, Juan), (apellido, Cancela), (fecha de ingreso, (2, enero, 1990)), (cédula de identidad, 5.789.231)}. Esto significa que esta entidad describe a un empleado llamado Juan Cancela con cédula de identidad número 5.789.231, que ingresó a la empresa el 2/1/1990 (ver figura 2).

Deben evitarse (o marcarse) los atributos derivables, esto es, atributos cuyos valores pueden calcularse a partir de otros atributos pertenecientes al mismo o a diferentes tipos de entidades. Por ejemplo un tipo de entidad FACTURA con atributos PRECIO UNITARIO y CANTIDAD, podría tener como atributo derivable IMPORTE TOTAL.

- d) El IDENTIFICADOR es un atributo o conjunto de atributos que permiten distinguir a cada entidad dentro del conjunto de entidades del mismo tipo. (ver más adelante CLAVES PRIMARIAS)

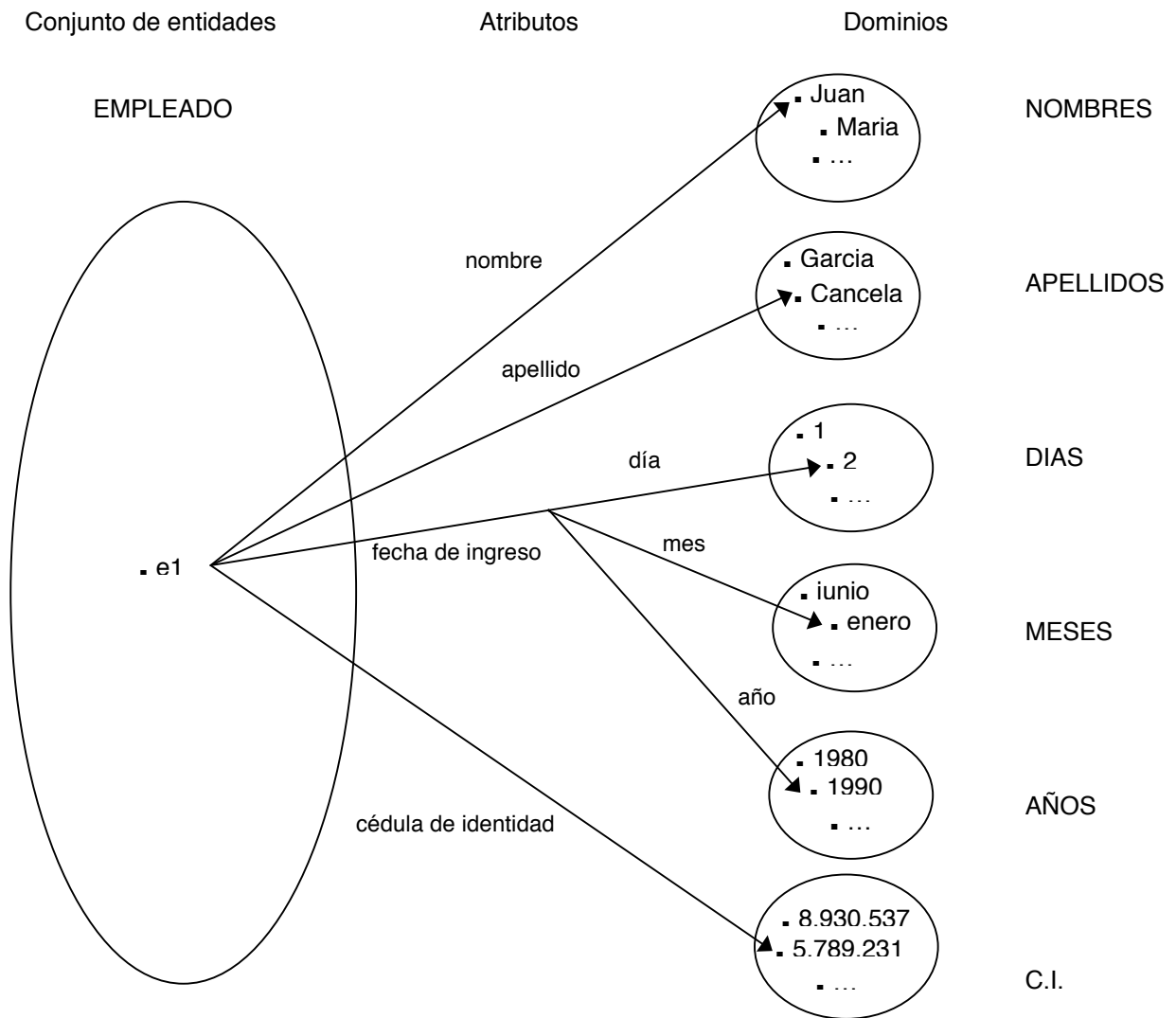


Fig. 2

INTERRELACIONES Y TIPOS DE INTERRELACIONES

Una *Interrelación* es una asociación entre dos o más entidades. Por ejemplo, la interrelación que asocia el cliente "Pérez" con la cuenta 259.

Un *Conjunto de Interrelaciones* (o *Tipo de Interrelación*) es un conjunto de interrelaciones del mismo tipo. Formalmente es una relación matemática sobre $n \geq 2$ conjuntos (no necesariamente disjuntos) de entidades. Si E_1, E_2, \dots, E_n son conjuntos de entidades, entonces el conjunto de interrelaciones R es un subconjunto del producto cartesiano

$$\{(e_1, e_2, \dots, e_n) / e_1 \in E_1, e_2 \in E_2, \dots, e_n \in E_n\},$$

donde la tupla (e_1, e_2, \dots, e_n) es una interrelación.

Consideremos los dos tipos de entidades **CLIENTE** y **CUENTA**. Definimos el tipo de interrelación **TENENCIA** que denomina la asociación entre los clientes y sus cuentas bancarias.

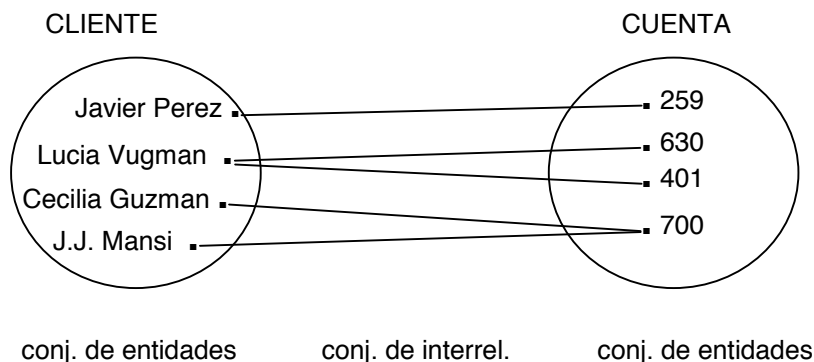
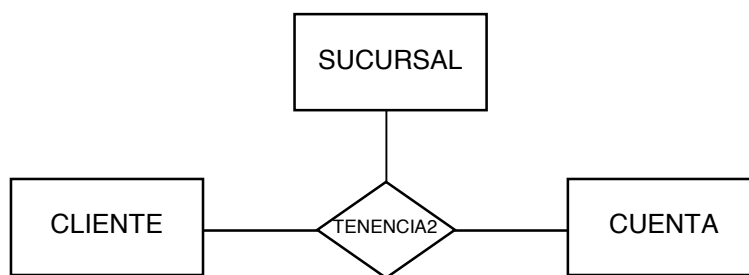


Fig. 3

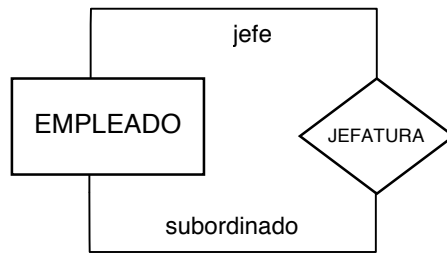


TENENCIA es un ejemplo de tipo de interrelación binario, esto es, un tipo de interrelación que involucra dos tipos de entidades.

Grado de un tipo de interrelación es el número de tipos de entidades que relaciona. TENENCIA es un tipo de interrelación de grado **dos**. Existen tipos de interrelación que involucran más de dos tipos de entidades. Como ejemplo, considérese la interrelación ternaria (Pérez, 259, Boedo) que especifica que el cliente Pérez tiene la cuenta 259 en la sucursal Boedo. Esta interrelación es una instancia del tipo de interrelación TENENCIA2 que involucra los tipos de entidades CLIENTE, CUENTA y SUCURSAL.



La función que una entidad desempeña en una interrelación se denomina su *rol*. Normalmente los roles están implícitos y no son especificados. Al hablar del tipo de interrelación TENENCIA sobreentendemos el rol que juega cada tipo de entidad: un CLIENTE "tiene" una CUENTA, una CUENTA "es (man)tenida por" un CLIENTE. Sin embargo, puede resultar útil especificar los roles para esclarecer el significado del tipo de interrelación. Tal es el caso cuando los tipos de entidades que participan en un tipo de interrelación no son distintos. Por ejemplo, las relaciones de subordinación/jefatura que existen entre los empleados de una empresa pueden representarse con el tipo de interrelación JEFATURA.



Un EMPLEADO "es jefe de" n EMPLEADOs, un EMPLEADO "es subordinado de" un EMPLEADO (su jefe). Cada interrelación del tipo JEFATURA puede modelizarse por un par ordenado de entidades EMPLEADO. Por ejemplo, el primer empleado del par toma el rol de subordinado, y el segundo toma el rol de jefe. De esta manera, todas las interrelaciones JEFATURA están caracterizados por pares (subordinado, jefe). Los pares (jefe, subordinado) están excluidos. JEFATURA es un tipo de interrelación grado **uno**.

También un tipo de interrelación puede tener atributos descriptivos. Por ejemplo, FECHA ULTIMO MOVIMIENTO puede ser un atributo del tipo de interrelación TENENCIA, indicando la última fecha en la que un cliente ha accedido su cuenta. La interrelación (Pérez, 259) estará descrita por {(fecha, 23 de mayo de 1998)} lo que significa que la ultima vez que Pérez accedió a la cuenta 259 fue el 23 de mayo de 1998.

DIAGRAMAS E-R

Un Modelo E-R consiste en una colección de tipos de entidades y tipos de interrelaciones más los atributos, dominios y restricciones necesarios para describir las propiedades que interesan de esos conjuntos. Una forma gráfica de representar un modelo de este tipo es mediante un diagrama E-R que utiliza la siguiente simbología:

- **rectángulos** que representan tipos de entidades
- **óvalos** que representan atributos
- **rombos** que representan tipos de interrelaciones
- **arcos** que asocian atributos a los tipos de entidades y tipos de entidades a tipos de interrelaciones

Cada componente está rotulado con su correspondiente nombre. Los roles se indican rotulando los arcos que conectan los rectángulos con los rombos.

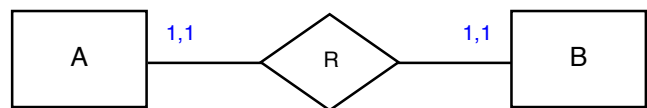
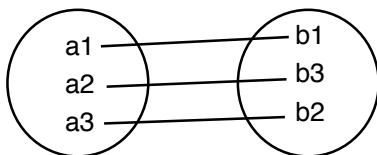
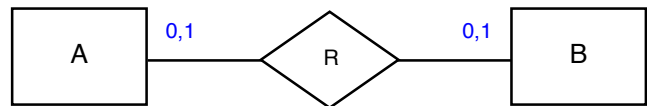
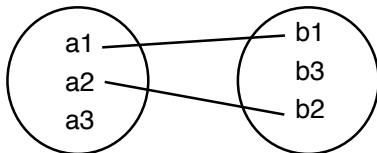
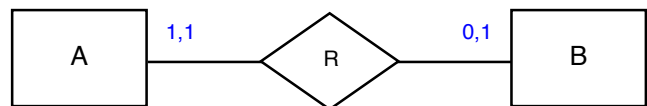
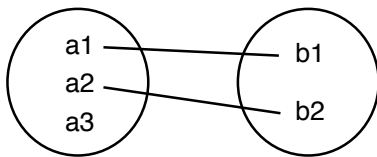
RESTRICCIONES DE CARDINALIDAD DE CORRESPONDENCIA

Las cardinalidades de correspondencia son restricciones importantes que expresan la cantidad (mínima y máxima) de entidades a las cuales puede asociarse una entidad a través de un tipo de interrelación. En esta sección nos ocuparemos de las cardinalidades de correspondencia entre tipos de interrelaciones binarias (es decir, interrelaciones entre dos tipos de entidades).

Para un tipo de interrelación binario **R** entre dos tipos de entidades **A** y **B**, la cardinalidad puede ser una de las siguientes:

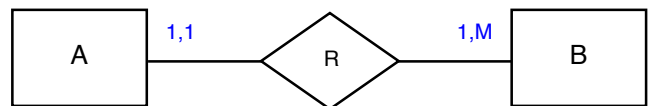
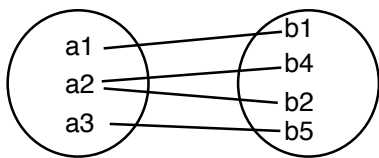
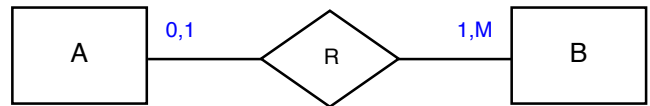
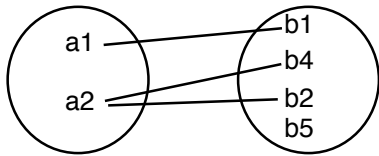
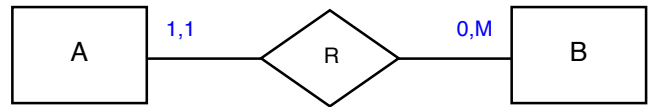
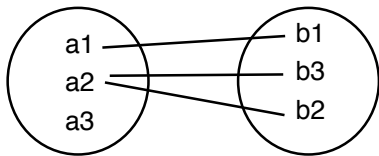
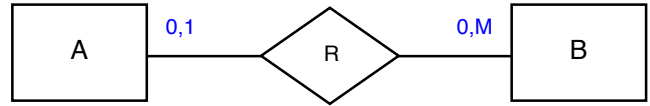
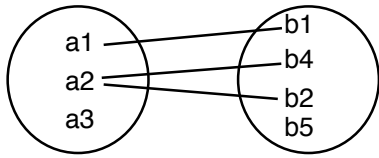
- **UNO A UNO (1:1)**

Una entidad en **A** está asociada como máximo a una entidad en **B**, y una entidad en **B** está asociada como máximo a una entidad en **A**.



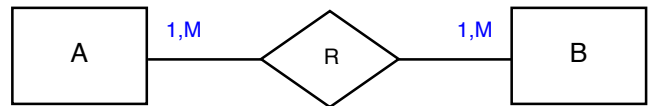
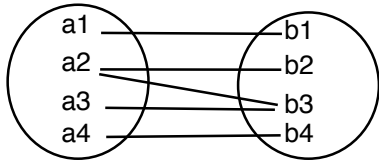
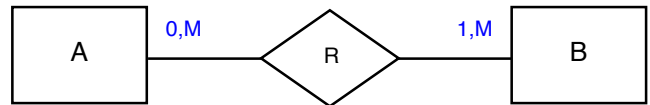
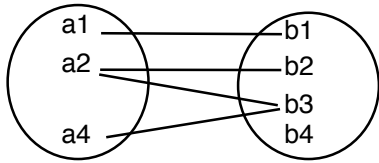
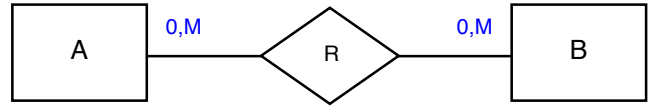
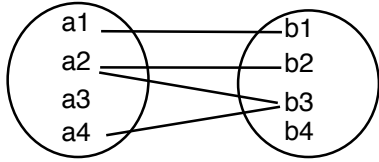
UNO A MUCHOS (1:M)

Una entidad en **A** está asociada como máximo con cualquier cantidad de entidades en **B**, y una entidad en **B** está asociada como máximo a una entidad en **A**.



▪ MUCHOS A MUCHOS (M:M)

Una entidad en **A** está asociada como máximo con cualquier cantidad de entidades en **B**, y una entidad en **B** está asociada como máximo con cualquier cantidad de entidades en **A**.



La cardinalidad de correspondencia apropiada para un conjunto de interrelaciones particular depende, obviamente, del mundo real que está siendo modelizado. Considérese el tipo de interrelación **TENENCIA**. Si en un banco en particular una cuenta puede pertenecer a un cliente solamente, y un cliente puede tener varias cuentas, entonces el tipo de interrelación es **uno-a-muchos** de CLIENTE a CUENTA.



Si una cuenta puede pertenecer a varios clientes en forma conjunta, el tipo de interrelación es **muchos-a-muchos**.

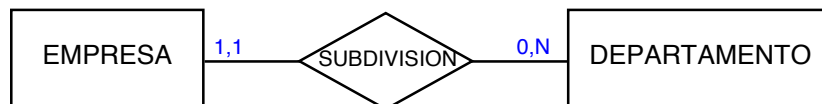


Veamos otros ejemplos:

- a) Todo empleado tiene asignado un escritorio para trabajar, en cada escritorio trabaja un solo empleado, no todos los escritorios tienen empleados asignados.



- b) Las empresas están divididas en departamentos, todo departamento pertenece a una empresa exclusivamente, no todas las empresas están formadas por departamentos.



- c) Todo empleado tiene habilidades, varios empleados pueden poseer la misma habilidad, pueden existir habilidades que no sean poseídas por ningún empleado.



RESTRICCIONES DE DEPENDENCIA EXISTENCIAL

Las dependencias de existencia conforman otra clase importante de restricciones. Específicamente, si la existencia de la entidad **S** depende de la existencia de la entidad **D**, entonces se dice que **S** tiene dependencia existencial o depende existencialmente de **D**. Operacionalmente, esto significa que si **D** es borrado, también **S** es borrado. La entidad **D** es la entidad *dominante* y **S** es la entidad *subordinada*.

Considérese los tipos de entidades CUENTA y TRANSACCION, y el tipo de interrelación MOVIMIENTO entre éstos, que especifica que para una cuenta particular pueden existir varias transacciones. Este tipo de interrelación es **uno-a-muchos** de CUENTA a TRANSACCION. Cada entidad TRANSACCION debe estar asociada a una entidad CUENTA. Si una entidad cuenta es borrada, entonces todas las entidades TRANSACCION asociadas también deben ser borradas. Sin embargo, las entidades TRANSACCION pueden ser borradas de la base de datos sin afectar ninguna entidad CUENTA. El tipo de entidad CUENTA, por lo tanto, es *dominante* y el tipo de entidad TRANSACCION es *subordinado* del tipo de interrelación MOVIMIENTO.



CLAVES PRIMARIAS - RESTRICCIONES DE IDENTIFICACION

Conceptualmente las entidades y interrelaciones individuales son distinguibles entre sí. Una **superclave** es un conjunto de uno a más atributos que, tomados colectivamente, permiten identificar unívocamente una entidad (o una interrelación) en el conjunto de entidades (o interrelaciones) del mismo tipo. Por ejemplo, el atributo NOMBRE DE CLIENTE del tipo de entidad CLIENTE no es una superclave, pues varios clientes pueden tener el mismo nombre. El atributo CEDULA DE IDENTIDAD del tipo de entidad CLIENTE es suficiente para distinguir una entidad CLIENTE de otra. Por lo tanto CEDULA DE IDENTIDAD es una superclave. Asimismo la combinación NOMBRE DE CLIENTE y CEDULA DE IDENTIDAD es una superclave para el tipo de entidad CLIENTE. Si **K** es una superclave, cualquier superconjunto de **K** también es una superclave.

Interesan generalmente las menores superclaves posibles. Esto es, estamos interesados en aquellas superclaves para las cuales ningún subconjunto propio es una superclave. Estas superclaves minimales se denominan **claves candidatas** o identificadores candidatos.

Vemos que aunque los atributos NOMBRE DE CLIENTE y CEDULA DE IDENTIDAD juntos son una superclave (pues permitirían distinguir entidades CLIENTE), su combinación no constituye una clave candidata porque el atributo CEDULA DE IDENTIDAD por sí solo es una superclave.

Es posible que varios conjuntos distintos de atributos puedan servir como claves candidatas. El atributo D.N.I., por ejemplo, podría ser suficiente para distinguir entre miembros del conjunto de entidades CLIENTE. Así, tanto {CEDULA DE IDENTIDAD} como {D.N.I.} son claves candidatas.

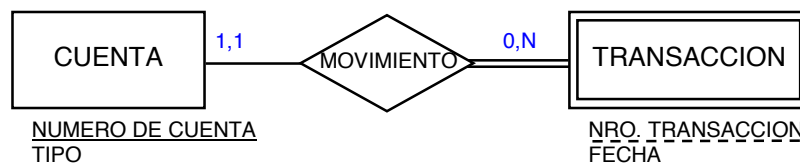
Usaremos el término **clave primaria** o identificador principal para denominar aquella clave candidata que es elegida, por el diseñador del modelo de datos, como medio principal para identificar entidades dentro de un conjunto de entidades.

Puede ocurrir que en un tipo de entidad ningún atributo (o conjunto de atributos) satisfaga las condiciones de identificador principal, esto es, permitir la identificación unívoca de una entidad dentro del conjunto. Por ejemplo, si para cada CUENTA las TRANSACCIONES se numeran secuencialmente desde 1, no será posible identificar por NRO.TRANSACCION una transacción dentro del conjunto de transacciones, diferentes transacciones sobre diferentes cuentas pueden tener el mismo número.

Se dice que una entidad tiene una *dependencia de identidad o identificación* si no es identificable por sus propios atributos. Este tipo de entidad que no tiene una clave primaria propia se denomina **entidad débil**, y normalmente está subordinada a otro tipo de entidad, que denominamos entidad fuerte, de la cual depende su identificación. Una entidad que tiene clave primaria propia es una **entidad fuerte**.

El concepto de entidades fuertes y débiles está relacionado con el concepto de *dependencia de existencia*. Una entidad débil es, por definición, una entidad *subordinada* a la entidad *dominante* de la cual depende su identidad. Toda *dependencia de identidad* es una *dependencia de existencia*, pero una *dependencia de existencia* no implica necesariamente una *dependencia de identidad*.

Un tipo de entidad débil se indica con un rectángulo doble. El tipo de interrelación que lo conecta con el tipo de entidad fuerte del cual depende se indica con arcos dobles.



Aunque un tipo de entidad débil no tiene clave primaria propia, es necesario distinguir, entre sí, las entidades débiles que dependen de una misma entidad fuerte. El **discriminador** de un tipo de entidad débil es el conjunto de atributos que permiten realizar esta distinción. Por ejemplo, el discriminador del tipo de entidad débil TRANSACCION es el atributo NRO. TRANSACCION, dado que para cada cuenta estos números de transacción identifican unívocamente una única transacción.

La clave primaria de un tipo de entidad débil está formada por la clave primaria de la entidad fuerte de la cual depende existencialmente, más su discriminador. En el caso del tipo de entidad TRANSACCION, su clave primaria es (NUMERO DE CUENTA, NRO. TRANSACCION), donde NUMERO DE CUENTA identifica la entidad dominante de una TRANSACCION, y NRO. TRANSACCION distingue las entidades TRANSACCION dentro de la misma cuenta.

Los tipos de interrelación también tienen claves primarias. En general, aunque no siempre, sus claves primarias están formadas por los atributos de las claves primarias de los tipos de entidad que asocian. Por ejemplo, CEDULA DE IDENTIDAD es la clave primaria de CLIENTE, y NUMERO DE CUENTA es la clave primaria de CUENTA. La clave primaria del tipo de interrelación TENENCIA es (CEDULA DE IDENTIDAD, NUMERO DE CUENTA). La clave primaria del tipo de interrelación MOVIMIENTO es (NUMERO DE CUENTA, NRO. TRANSACCION). Obsérvese que en este último caso la clave primaria del tipo de interrelación coincide, trivialmente, con la del tipo de entidad débil; en general esto sucede con cada tipo de interrelación binario *funcional* (cardinalidades 1:M o 1:1), en que la condición de *minimalidad* determina que su clave primaria coincida con la de uno de los tipos de entidad que asocia, según el sentido de la *funcionalidad* de la relación (se verán ejemplos en el Caso de Estudio).

Puede darse el caso de un tipo de interrelación que sea identificado por un atributo propio.

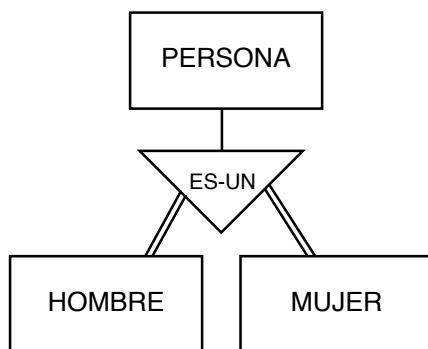
GENERALIZACION Y ESPECIALIZACION

La generalización y la especialización son dos tipos de interrelaciones con características particulares.

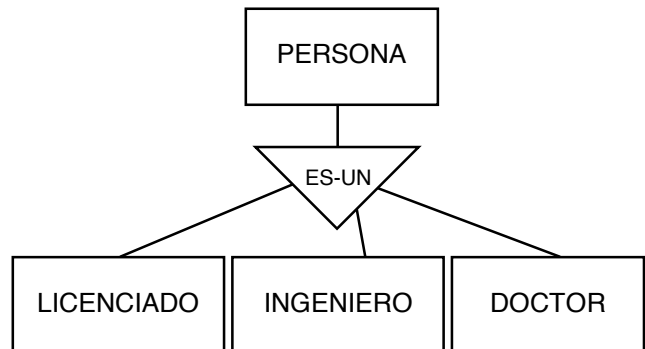
Generalización es el resultado de unir dos o más conjuntos disjuntos de entidades de menor nivel (subtipos) para producir un conjunto de entidades de mayor nivel (supertipo). En una generalización toda entidad del supertipo debe ser una entidad de 1 y sólo 1 subtipo.

Especialización es el resultado de tomar uno o varios subconjuntos de un conjunto de entidades de mayor nivel (supertipo) para formar conjuntos de entidades de menor nivel (subtipos). En una especialización una entidad del supertipo puede pertenecer a 0, 1, o N subtipos.

En un diagrama E-R tanto la generalización como la especialización se representan a través de un triángulo ES-UN (ISA). La generalización se distingue de la especialización mediante arcos dobles entre el triángulo y los subtipos de entidades.



Generalización



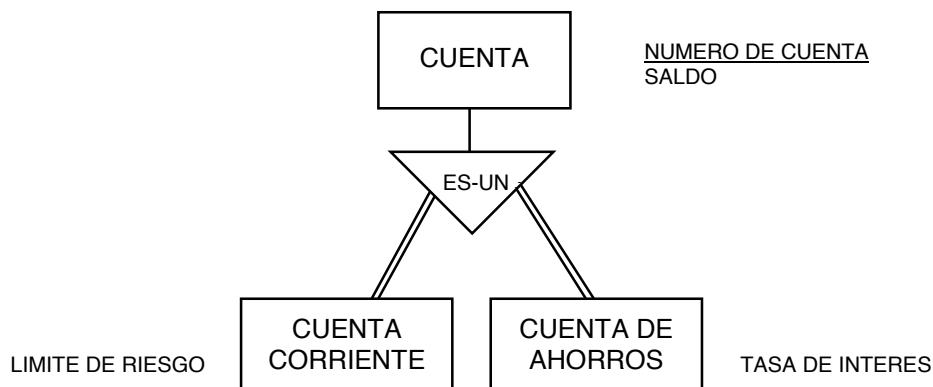
Especialización

Las cardinalidades de estos tipos de interrelación se omiten en el diagrama porque por definición son siempre las mismas: 1 a 1 del supertipo a cada uno de los subtipos.

Cada uno de los subtipos de entidades puede tener atributos propios. Además de estos atributos propios, tanto en la generalización como en la especialización, los atributos de las entidades de mayor nivel son **heredados** por las entidades de menor nivel.

Veamos otros ejemplos:

- a) Podemos clasificar las cuentas de un banco en cuentas corrientes y cajas de ahorro. Toda cuenta tiene un número de cuenta y un saldo. Las cuentas de ahorro están adicionalmente descriptas por una tasa de interés y las cuentas corrientes por un límite de riesgo.

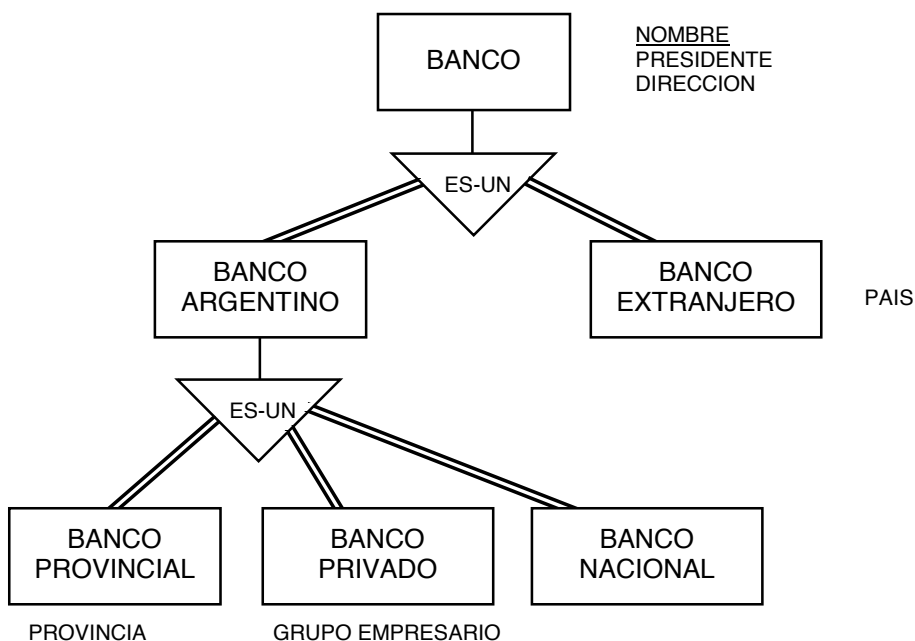


Por qué una generalización y no una especialización?

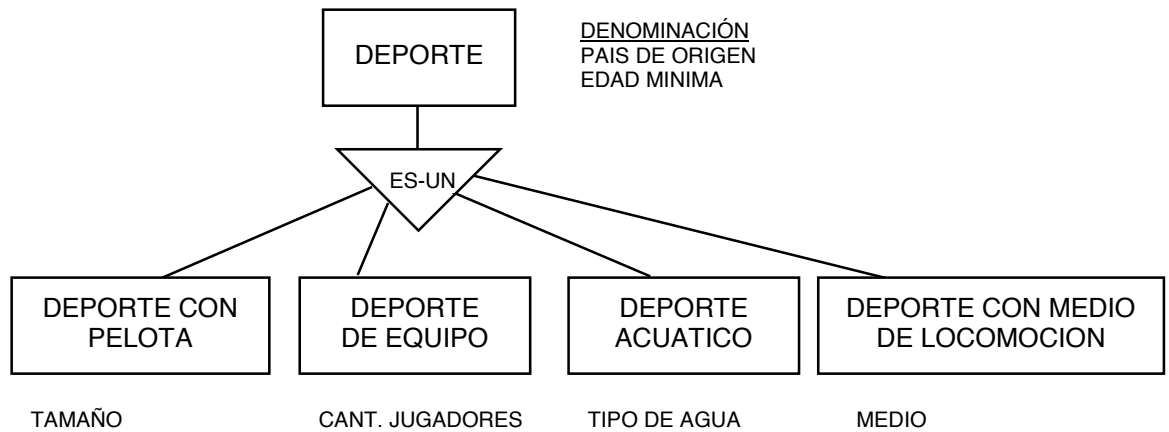
Existen similitudes entre el tipo de entidad **CUENTA CORRIENTE** y el tipo de entidad **CUENTA DE AHORRO** en el sentido que tienen varios atributos en común. Una generalización requiere que toda entidad **CUENTA** sea una entidad **CUENTA DE AHORRO** o (excluyente) una entidad **CUENTA CORRIENTE**.

Cada entidad del subtipo está descripta por un conjunto de atributos que incluyen todos los atributos del tipo de entidad **CUENTA** más los atributos propios del subtipo. Es decir, una **CUENTA DE AHORRO** está descripta por los atributos **NUMERO DE CUENTA**, **SALDO** y **TASA DE INTERES**. Una **CUENTA CORRIENTE** está descripta por los atributos **NUMERO DE CUENTA**, **SALDO** y **LIMITE DE RIESGO**.

- b) Considerando la plaza bancaria argentina, encontramos bancos nacionales, provinciales y privados, todos éstos argentinos; y bancos extranjeros. Podríamos generar el siguiente modelo:



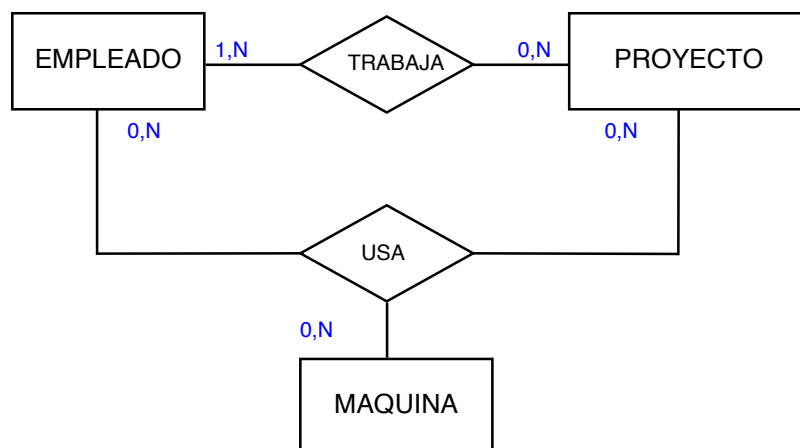
- c) Hablando de deportes podemos querer guardar distinto tipo de información de acuerdo a los distintos tipos de deportes. Cada uno de los subtipos (DEPORTE CON PELOTA, DEPORTE DE EQUIPO, etc.) es un subconjunto del conjunto de todos los deportes (DEPORTE), pero no son excluyentes el uno del otro, incluso pueden existir deportes que no pertenecen a ninguno de estos subconjuntos.



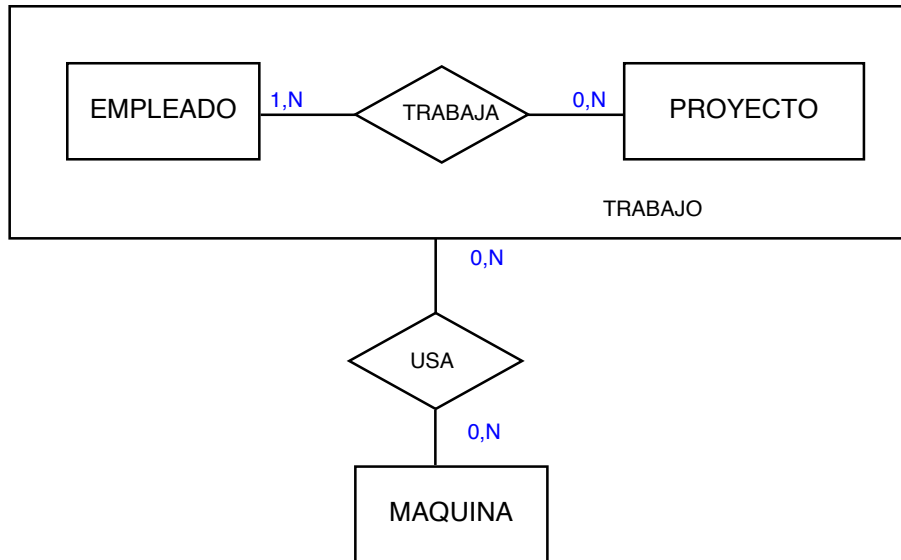
AGREGACION

Una limitación del modelo E-R es la imposibilidad de expresar interrelaciones entre interrelaciones. La solución es usar la [agregación](#). La agregación es una abstracción que permite manejar interrelaciones como entidades de mayor nivel.

Considérese el siguiente caso: algunos empleados trabajan en proyectos y durante su trabajo usan diferentes máquinas. Usando las construcciones conocidas hasta ahora del modelo E-R, obtendríamos el siguiente diagrama:



Sin embargo, este diagrama no muestra claramente el hecho de que una máquina es usada por "un EMPLEADO asignado a un PROYECTO". Para ello, podemos realizar la agregación del tipo de interrelación TRABAJA y los tipos de entidades EMPLEADO y PROYECTO en un tipo de entidad de mayor nivel llamado TRABAJO. Este tipo de entidad puede ser tratado como cualquier otro tipo de entidad.



CASO DE ESTUDIO

Sea una empresa de ferrocarriles destinada a prestar servicios de transporte ferroviario en distintos países. Los países están compuestos por provincias.

Los empleados de la empresa que no sean directivos son asignados a talleres (mecánicos, técnicos, etc.) o trenes (conductores, inspectores, etc.). Se asumirá que existen empleados dirigidos por otros.

Existen distintos tipos de trenes los cuales son mantenidos en talleres; éstos han de mantener al menos un tipo de tren.

A cada tren se le asigna una ruta, la cual pasa por un conjunto de estaciones, y que está compuesta por tramos para los cuales existen plazas (asientos disponibles) ².

Se admite que existan ciudades de igual nombre en distintas provincias, como así también que existan nombres de provincias iguales en países distintos.

Las estaciones están localizadas en ciudades al igual que los talleres, con la observación que estos últimos deben encontrarse en las inmediaciones de alguna estación.

Distinguiremos seis etapas en el diseño de un modelo E-R:

1. Identificar y definir los **tipos de entidades** que nos interesan
2. Identificar y definir los **tipos de interrelaciones** que nos interesan
3. Definir los **atributos** que nos interesan (y sus dominios) para los **tipos de entidades**
4. Definir los **atributos** que nos interesan (y sus dominios) para los **tipos de interrelaciones**
5. Determinar las **claves primarias** de los tipos de entidades
6. Determinar las **claves primarias** de los tipos de interrelaciones

² Al efecto de simplificar la solución se establecen las siguientes hipótesis:

- Se consideran los recorridos en un solo sentido (dirección)
- Se considera que un tren no puede realizar mas de un viaje por día.

Definición de tipos de entidades

PAIS	División política del territorio por donde transitan los TRENes
PROVINCIA	Subdivisión administrativa de los PAISes
CIUDAD	Todo lugar perteneciente a alguna PROVINCIA donde al menos exista una ESTACION ferroviaria
ESTACION	Lugar dentro de una CIUDAD donde pueden operar los TRENes
TALLER	Lugar donde los TRENes son atendidos tanto para mantenimiento preventivo como correctivo
TRAMO	Recorrido ferroviario entre ESTACIONes, sin paradas intermedias
ruta	Conjunto de TRAMOs consecutivos para ir desde la ESTACION origen hasta la ESTACION terminal de esa RUTA
TREN	Medio de transporte en el cual se ejecuta un recorrido
TIPO-TREN	Categorías en que son clasificados los TRENes
FECHA	Día, mes y año en que esta programado un viaje
EMPLEADO	Personal que trabaja en la empresa ferroviaria

Definición de tipos de interrelaciones

"Los países esta compuestos por provincias."



"Se admite que existan ciudades de igual nombre en distintas provincias, como así también que existan nombres de provincias iguales en países distintos."



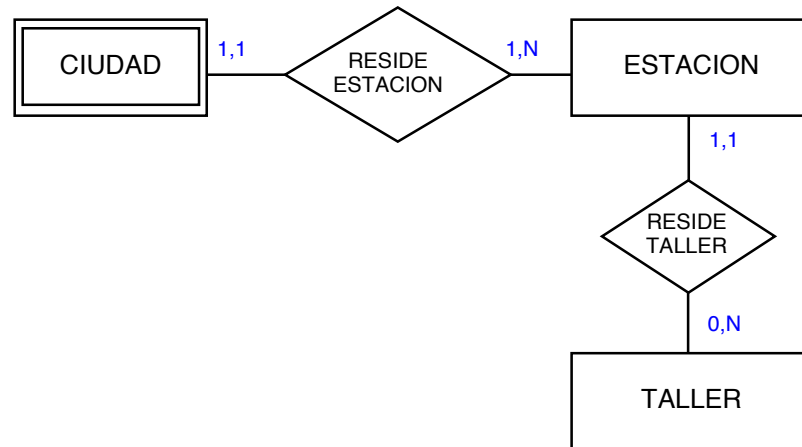
"Existen distintos tipos de trenes...."



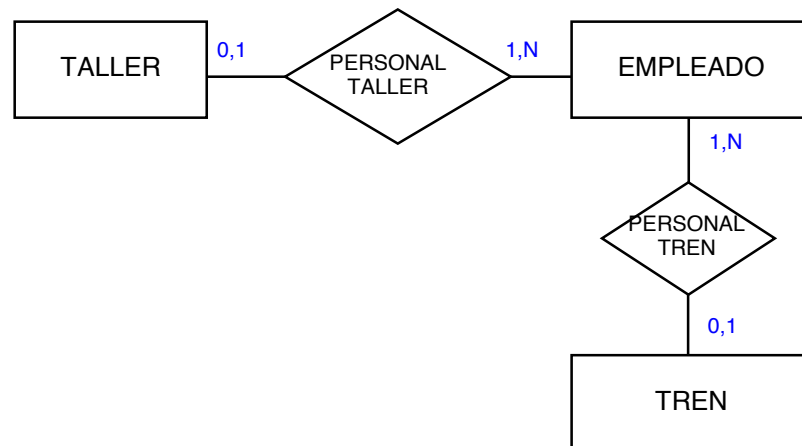
"...tipos de trenes los cuales son mantenidos en talleres; éstos han de mantener al menos a un tipo de tren."



"Las estaciones están localizadas en ciudades al igual que los talleres, con la observación que estos últimos deben encontrarse en las inmediaciones de alguna estación."

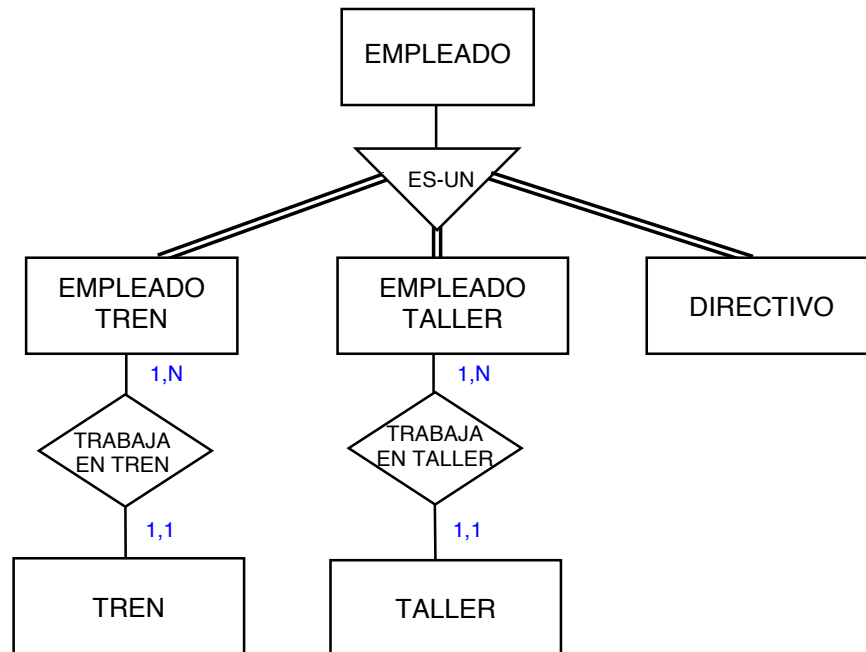


"Los empleados de la empresa que no sean directivos son asignados a talleres (mecánicos, técnicos, etc.) o trenes (conductores, inspectores, etc.)"

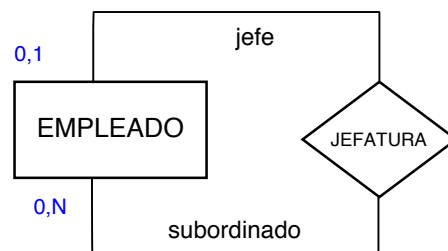


Con esta representación no estamos recogiendo la restricción de exclusión contenida en la afirmación de que todo empleado o es directivo o es de tren o es de taller. Si observamos con detenimiento el tipo de entidad EMPLEADO podemos apreciar que en su relación de trabajo se particiona en tres tipos de entidades (o subtipos de entidades): los que trabajan en talleres, los que trabajan en trenes y los directivos. Esta partición nos está señalando que son de interés los tipos de entidad DIRECTIVO, EMPLEADO TALLER y EMPLEADO TREN, que no habíamos detectado en el paso 1 (Definición de tipos de entidades), al cual deberíamos agregar las siguientes definiciones:

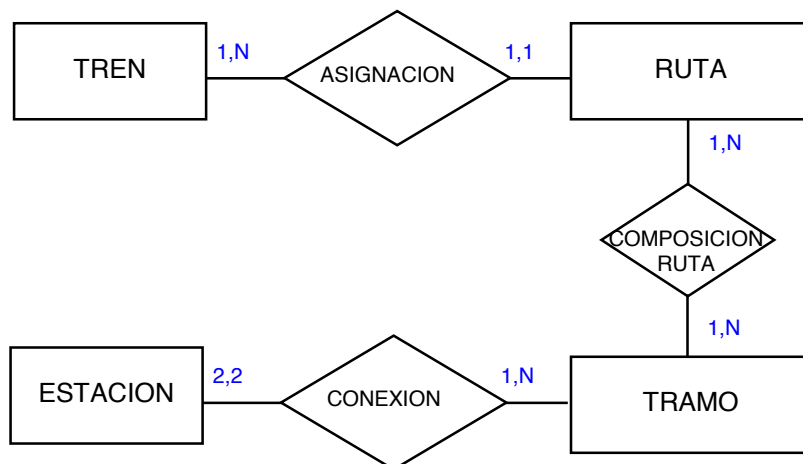
EMPLEADO TALLER EMPLEADO asignado para trabajar en un TALLER
 EMPLEADO TREN EMPLEADO asignado para trabajar en un TREN
 DIRECTIVO EMPLEADO que cumple funciones directivas



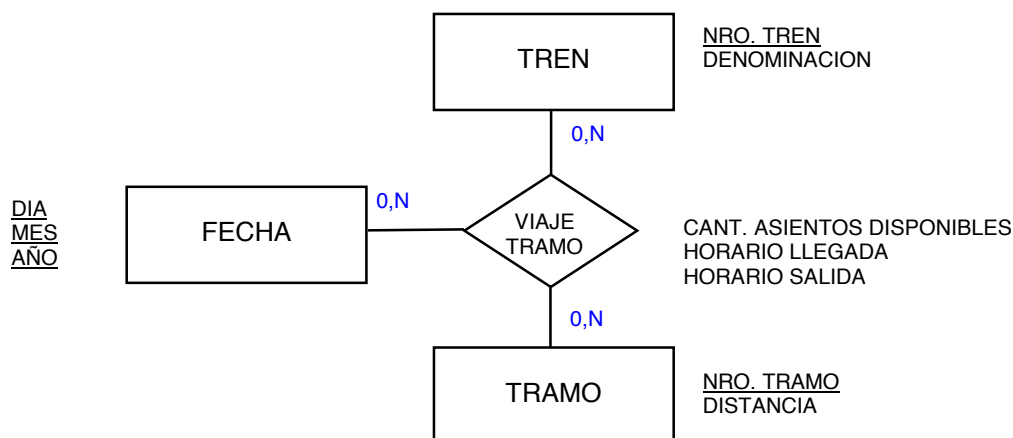
"Se asumirá que existen empleados dirigidos por otros."



"A cada tren se le asigna una ruta, la cual pasa por un conjunto de estaciones, y que está compuesta por tramos...."



"... tramos.. para los cuales existen plazas (asientos disponibles). "



Para determinar las cardinalidades de este tipo de interrelación ternaria (intervienen tres tipos de entidades) procedemos de la siguiente forma:

- Cada par TRAMO-FECHA se asocia con 0 a N TRENEs (es decir, en una determinada FECHA y por un determinado TRAMO, pasan 0 a N TRENEs).
- Cada par FECHA-TREN se asocia con 0 a N TRAMOs (es decir, en una determinada FECHA un TREN particular, pasa por 0 a N TRAMOs).
- Cada par TRAMO-TREN se asocia con 0 a N FECHAs (es decir, un determinado TREN pasa por un determinado TRAMO, en 0 a N FECHAs).

En cuanto a la identificación de tipos interrelaciones realizada, la definición de cada uno de los tipos de interrelaciones del modelo propuesto es la siguiente:

COMPOSICION PAIS	Composición de países en provincias (dentro de un país los nombres de las provincias son únicos).
RESIDE CIUDAD	Composición de provincias en ciudades (dentro de una provincia los nombres de las ciudades son únicos).
CLASIFICACION	Clasificación de trenes según su tipo.
MANTENIMIENTO	Mantenimiento de los trenes por talleres, según el tipo de tren.
RESIDE ESTACION	Existencia de una estación en una ciudad.
RESIDE TALLER	Existencia de un taller en una estación.
TRABAJA EN TREN	Asignación de empleados a los trenes.
TRABAJA EN TALLER	Asignación de empleados a los talleres.
JEFATURA	Relación jerárquica ente empleados y sus jefes.
ASIGNACION	Asignación de trenes a sus rutas.
COMPOSICION RUTA	Composición de rutas por tramos.
CONEXION	Conexión de tramos con dos estaciones.
VIAJE TRAMO	Recorrido de un tren en un tramo en una fecha determinada.

Identificación de atributos para tipos de entidades

PAIS	(nombre_país)
PROVINCIA	(nombre_provincia)
CIUDAD	(nombre_ciudad)
ESTACION	(nombre_estación, altura)
TALLER	(nro_taller, dirección)
TRAMO	(nro_tramo, distancia)
RUTA	(nro_ruta)
TREN	(nro_tren, denominación)
TIPO-TREN	(nombre_tipo, capacidad)
FECHA	(día, mes, año)
EMPLEADO	(nro_empleado, nombre_empleado, dirección_empleado)
EMPLEADO TALLER	(especialidad)
EMPLEADO TREN	(jornada, tipo_trabajo)
DIRECTIVO	(profesión, cargo)

Identificación de atributos para tipos de interrelaciones

VIAJE TRAMO	(cant_asientos_disponibles, horario_llegada, horario_salida)
-------------	--

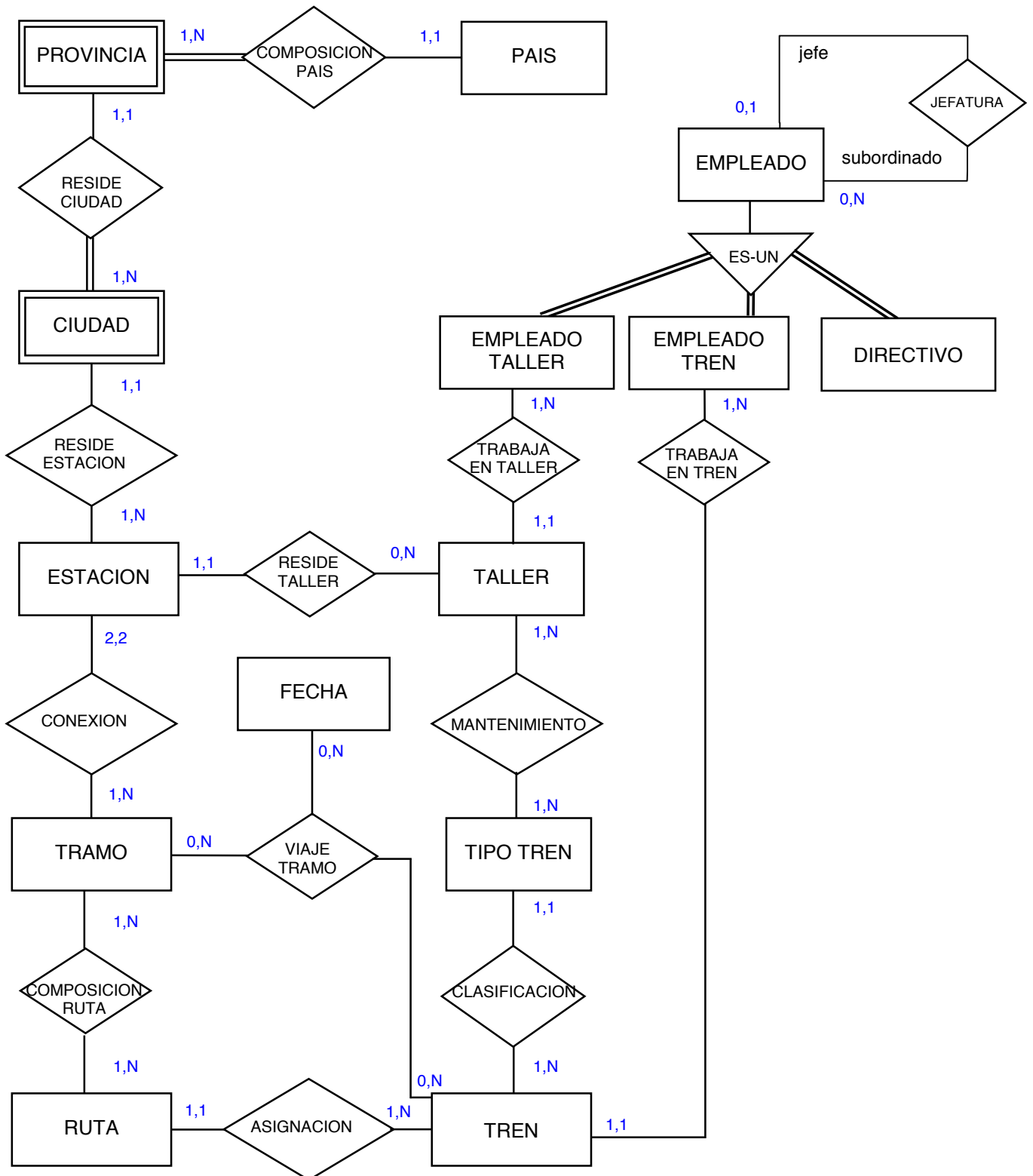
Identificación de claves primarias para tipos de entidades

PAIS	(nombre_país)
PROVINCIA	(nombre_país, nombre_provincia)
CIUDAD	(nombre_país, nombre_provincia, nombre_ciudad)
ESTACION	(nombre_estación)
TALLER	(nro_taller)
TRAMO	(nro_tramo)
RUTA	(nro_ruta)
TREN	(nro_tren)
TIPO-TREN	(nombre_tipo)
FECHA	(día, mes, año)
EMPLEADO	(nro_empleado)
EMPLEADO TALLER	(nro_empleado)
EMPLEADO TREN	(nro_empleado)
DIRECTIVO	(nro_empleado)

Identificación de claves primarias para tipos de interrelaciones

COMPOSICION PAIS	(nombre_país, nombre_provincia)
RESIDE CIUDAD	(nombre_país, nombre_provincia, nombre_ciudad)
CLASIFICACION	(nro_tren)
MANTENIMIENTO	(nombre_tipo, nro_taller)
RESIDE ESTACION	(nombre_estación)
RESIDE TALLER	(nro_taller)
TRABAJA EN TREN	(nro_empleado)
TRABAJA EN TALLER	(nro_empleado)
JEFATURA	(nro_empleado_jefe, nro_empleado_subordinado)
ASIGNACION	(nro_tren)
COMPOSICION RUTA	(nro_ruta, nro_tramo)
CONEXION	(nro_tramo, nombre_estación)
VIAJE TRAMO	(día, mes, año, nro_tren, nro_tramo)

Diagrama ENTIDAD-INTERRELACION



MODELO DE DATOS RELACIONAL

INTRODUCCION

El Modelo de Datos Relacional fue propuesto por E. F. Codd en 1970. La mayoría de los sistemas de bases de datos comercializados hasta 1980 utilizaban el Modelo Red o el Jerárquico, pero estos Modelos de Datos fueron siendo desplazados gradualmente por el Relacional debido a su sencillez y flexibilidad.

El Modelo Relacional, basado en conceptos muy simples, tiene asociada una teoría que no puede ser separada del mismo: **la teoría de la Normalización de las Relaciones.**

Esta teoría tiene por objeto suprimir los datos redundantes, eliminar los comportamientos anormales durante las actualizaciones y facilitar la comprensión de las relaciones semánticas entre los datos.

Todo Modelo de Datos se caracteriza por tres conjuntos:

- 1) ESTRUCTURAS
- 2) RESTRICCIONES (implícitas y explícitas)
- 3) OPERACIONES

En el Modelo RELACIONAL esta caracterización se manifiesta de la siguiente forma:

- | | |
|-----------------------------|------------------------------------|
| 1) ESTRUCTURAS | ⇒ dominios atómicos |
| | ⇒ relaciones |
| 2) RESTRICCIONES IMPLICITAS | ⇒ regla de integridad de entidad |
| | ⇒ regla de integridad referencial |
| 3) OPERACIONES | ⇒ álgebra relacional o equivalente |

CONCEPTOS BASICOS DEL MODELO

DOMINIO

Un dominio es un conjunto homogéneo de valores.

Ejemplos:

El dominio de los números enteros, el dominio de los números reales, el dominio de las vocales, etc.

$$V = \{a, e, i, o, u\}$$

RELACION

El concepto "relación" deriva directamente de la teoría de conjuntos. **Es un subconjunto del producto cartesiano de una lista de dominios. Está caracterizada generalmente por un nombre.**

Recordemos que el producto cartesiano de un conjunto de dominios $D_1, D_2, D_3, \dots, D_n$ es el conjunto de las tuplas (v_1, v_2, \dots, v_n) tales que v_i pertenece a D_i .

Ejemplo:

El producto cartesiano de los dominios:

$$D1 = \{\text{azul, blanco, rojo}\}$$

$$D2 = \{0, 1\}$$

está formado por las seis tuplas

$$D1 \times D2 = \{(\text{azul}, 0), (\text{azul}, 1), (\text{blanco}, 0), (\text{blanco}, 1), (\text{rojo}, 0), (\text{rojo}, 1)\}$$

A partir de $D1 \times D2$ podemos componer una relación r

$$r = \{(azul,0),(azul,1),(blanco,0)\}$$

De una forma más sencilla, podemos considerar una relación como una tabla de dos dimensiones en que las columnas corresponden a los dominios y las filas contienen las tuplas:

D1	D2
azul	0
azul	1
blanco	0

Dominio D1

Tupla (blanco,0)

ATRIBUTO

Los matemáticos definen una relación como un subconjunto del producto cartesiano de una lista de dominios, usando el entero **1** para denotar el atributo cuyo dominio aparece en el primer lugar en la lista de dominios: el entero **2** para el atributo cuyo dominio aparece en segundo lugar, etc.

El Modelo Relacional, en cambio, otorga un nombre a cada columna sin conferir importancia alguna al orden en el que se encuentran las mismas.

Sea una relación r conteniendo n tuplas. Sea t una tupla de la relación. Se utilizará la notación $t[A]$ para denotar el valor en t del atributo (o conjunto de atributos) A .

$$t[A] = \text{valor en la tupla } t \text{ del atributo } A$$

Adicionalmente se utilizará la notación $r.A$ para referirse al atributo A perteneciente a r .

$$r.A = \text{atributo } A/A \in r$$

ESQUEMA DE RELACION

En un primer enfoque, un **esquema de relación** es el nombre de la relación seguido de la lista de los atributos con sus dominios.

Debe diferenciarse entre el *esquema de relación* y las *instancias* de la relación: el primero describe la intención, mientras que una instancia de la relación describe su extensión en un momento dado.

Se adoptará la convención de usar minúsculas para las relaciones y mayúsculas para los esquemas de relación.

Ejemplo:

Dada una relación *automóvil* que incluye los atributos *número de matrícula*, *marca*, *modelo*, y *color*, su esquema es:

AUTOMOVIL = (NM: entero,
MARCA: char(10),
MODELO: char(4),
COLOR: color)

El hecho de que *automóvil* es una relación con el esquema AUTOMOVIL se expresa al escribir:

automóvil (AUTOMOVIL)

SUPERCLAVE, CLAVE CANDIDATA, CLAVE PRIMARIA, CLAVE FORÁNEA

Los conceptos *superclave*, *clave candidata* y *clave primaria* que se explicaron para el Modelo Entidad-Interrelación son aplicables al Modelo Relacional.

SUPERCLAVE

Es todo conjunto de uno o más atributos pertenecientes al esquema de relación **R** que permita identificar en forma unívoca a cada tupla **t** de la relación **r(R)**.

CLAVE CANDIDATA: UNICIDAD - MINIMALIDAD

Superclave para la cual ningún subconjunto propio es una superclave. En términos formales:

Sea **R** un esquema de relación con atributos A_1, A_2, \dots, A_n . El conjunto de atributos $K = (A_1, A_2, \dots, A_k)$ es una *clave candidata* de **R** si, y sólo si, satisface estas dos propiedades independientes del tiempo:

UNICIDAD

En todo momento, no existen dos tuplas distintas de **r(R)** que tengan el mismo valor para A_1 , el mismo valor para A_j , ..., el mismo valor para A_k .

MINIMALIDAD

Ningún atributo A_1, A_2, \dots, A_r puede ser eliminado del conjunto **K** sin que pierda la propiedad de unicidad.

CLAVE PRIMARIA

Se utiliza el término *clave primaria* para referir a la *clave candidata* que elija el diseñador de la base de datos como la forma principal de identificar a las tuplas dentro de una relación

CLAVE FORÁNEA

En una relación **r(R)** una *clave foránea* es un atributo, o combinación de atributos, que es *clave primaria* en otra relación **s(S)**.

Nótese que la *clave primaria* y la *clave foránea* deben estar definidas sobre el mismo dominio.

Los esquemas de relación **R** y **S** no son necesariamente distintos.

La relación **r(R)** se denomina *relación referenciante* y la relación **s(S)** *relación referenciada*.

Las *claves foráneas* representan las referencias de una relación a otra (es por esta razón que en algunas propuestas de Diagrama de Estructuras de Información basadas en el Diagrama Entidad-Interrelación surge lo que se denomina *atributo referencial*, para denotar, precisamente, una clave foránea del Modelo Relacional).

MODELO RELACIONAL: RESTRICCIONES

El Modelo Relacional tiene dos restricciones inherentes o implícitas, llamadas *reglas de integridad*:

- 1) REGLA DE INTEGRIDAD DE ENTIDAD
- 2) REGLA DE INTEGRIDAD REFERENCIAL

que refieren a las claves primarias y a las claves foráneas, respectivamente.

REGLA DE INTEGRIDAD DE ENTIDAD

Ningún atributo de la clave primaria de un esquema de relación puede tomar valor nulo.

Valor nulo significa que la información no se conoce o que la propiedad que describe el atributo no es aplicable.

REGLA DE INTEGRIDAD REFERENCIAL

Si la relación $r(R)$ incluye una clave foránea F que es la clave primaria P de una relación $s(S)$, entonces todo valor de F en $r(R)$ debe ser totalmente nulo o ser igual al valor de P en alguna tupla de $s(S)$.

R y S no son necesariamente distintos.

ALGEBRA RELACIONAL

El álgebra relacional es un lenguaje de consulta procedimental. Existen cinco operaciones básicas que son:

Operaciones unarias:

1. Selección (o Restricción)
2. Proyección

Operaciones binarias:

3. Unión
4. Diferencia
5. Producto cartesiano

Todas ellas producen como resultado una nueva relación.

Además de las cinco operaciones fundamentales, introduciremos otras operaciones, a saber

Operaciones binarias:

6. Intersección
7. Join
8. División

Estas se definirán en términos de las operaciones fundamentales.

SELECCIÓN o RESTRICCIÓN

La selección de la relación $r(R)$ por un predicado (o criterio de selección) P , es una relación $r'(R)$, con el mismo esquema R , en la que las tuplas son aquellas pertenecientes a r que satisfacen el predicado P .

Se utiliza la siguiente notación para la operación Selección aplicada sobre una relación r con un predicado P :

$$\sigma_P(r)$$

Para clarificar esta definición, introduzcamos el concepto de *predicado*.

Un *predicado atómico* es una expresión lógica simple formada por un nombre de atributo ligado a una constante, o a otro nombre de atributo, mediante un operador de comparación θ , tal que θ pertenece a $\{ <, >, <=, >=, =, <> \}$. Un predicado atómico es, pues, del tipo $A_i \theta C$, donde A_i es un nombre de atributo y C es una constante u otro nombre de atributo.

Los *predicados atómicos* se combinan mediante *conectores lógicos* conjuntivos y/o disyuntivos para formar *predicados compuestos*.

Conectores:

- conjunción (operador AND): \wedge
- disyunción inclusiva (operador OR): \vee

Operadores:

- Negación: \neg

En general, un predicado contendrá expresiones del tipo:

$$P = (A_1 \theta C_1) \wedge ((A_2 \theta C_2) \vee \neg(A_3 \theta C_3))$$

Ejemplo de Selección:

Dada la relación **automóvil(AUTOMOVIL)** :

AUTOMOVIL	NM	MARCA	MODELO	COLOR
	SQY687	Chevrolet	Corsa 4p.	Verde
	AQA121	Audi	A3	Gris perla
	BEX369	Chrysler	Grand Cherokee	Verde
	RCP333	Renault	Megane	Azul océano
	UTU025	Renault	Clío	Azul océano

fig.2

La operación $\sigma_{\text{COLOR} = \text{"Verde"}}(\text{automóvil})$ tiene como relación resultante a:

AUTOMOVIL	NM	MARCA	MODELO	COLOR
	SQY687	Chevrolet	Corsa 4p.	Verde
	BEX369	Chrysler	Grand Cherokee	Verde

fig.3

PROYECCION

La proyección de una relación r , de esquema $R=(A_1, A_2, \dots, A_n)$, sobre los atributos $A_{i1}, A_{i2}, \dots, A_{ip}$ (donde se cumple $I_j < > I_k$ y $p < n$) es una relación r' , de esquema $R'=(A_{i1}, A_{i2}, \dots, A_{ip})$, cuyas tuplas se obtienen por eliminación en r de los atributos de R que no pertenecen a R' y por supresión de las tuplas duplicadas.

Se utiliza la siguiente notación para la operación Proyección de una relación r sobre los atributos:

$$\pi_{A_{i1}, A_{i2}, \dots, A_{ip}}(r)$$

Ejemplo de Proyección:

Dada la relación **automóvil(AUTOMOVIL)** de la fig.2, la operación $\pi_{MARCA, COLOR}(r)$ tiene como relación resultante a:

MARCA	COLOR
Chevrolet	Verde
Audi	Gris perla
Chrysler	Verde
Renault	Azul océano

fig 4.

Nótese que se han eliminado las columnas correspondientes a los atributos NM y MODELO. Al eliminar estas dos columnas quedaría repetida la tupla (Renault, Azul océano), por lo tanto, para cumplir con la definición de relación, se elimina la tupla repetida.

UNION

La unión de dos relaciones **r1** y **r2** que tienen el mismo esquema, es una relación **r** con igual esquema, conteniendo el conjunto de tuplas que pertenecen a **r1** o a **r2** o a ambas relaciones.

El requisito que deben cumplir los esquemas de relación de ambos operandos, se sintetiza diciendo que **r1** y **r2** son **UNIONCOMPATIBLES**.

Se ha adoptado la siguiente notación para la operación unión entre dos relaciones **r1** y **r2**:

r1 U r2

Ejemplo de Unión:

Dadas las relaciones **auto1(AUTOMOVIL)** y **auto2(AUTOMOVIL)**:

auto1

AUTOMOVIL	NM	MARCA	MODELO	COLOR
	CSW698	Chevrolet	Corsa 4p.	Verde
	BMW666	Audi	A3	Negro
	AXA123	Renault	Megane	Blanco

auto2

AUTOMOVIL	NM	MARCA	MODELO	COLOR
	PEZ888	Honda	Accord	Gris Perla
	REG444	Mitsubishi	Colt	Verde
	BMW666	Audi	A3	Negro

fig.5

La relación resultante de la operación **auto1 U auto2** es:

AUTOMOVIL	NM	MARCA	MODELO	COLOR
	CSW698	Chevrolet	Corsa 4p.	Verde
	AXA123	Renault	Megane	Blanco
	PEZ888	Honda	Accord	Gris Perla
	REG444	Mitsubishi	Colt	Verde
	BMW666	Audi	A3	Negro

fig.6

DIFERENCIA

La diferencia de dos relaciones **r1** y **r2**, ambas con el mismo esquema y en el orden **r1 - r2**, es una relación **r** con igual esquema, conteniendo las tuplas que pertenecen a **r1** y no pertenecen a **r2**.

Nótese que **r1** y **r2** deben ser *UNIONCOMPATIBLES*.

Se utiliza la siguiente notación:

r1 - r2

Ejemplo de Diferencia:

Dadas las relaciones **auto1(AUTOMOVIL)** y **auto2(AUTOMOVIL)** presentes en la figura 5, la relación resultante de la operación **auto1 - auto2** es:

AUTOMOVIL	NM	MARCA	MODELO	COLOR
	CSW698	Chevrolet	Corsa 4p.	Verde
	AXA123	Renault	Megane	Blanco

fig.7

PRODUCTO CARTESIANO

El producto cartesiano de dos relaciones **r1(R1)** y **r2(R2)** es una relación **r** que tiene por esquema la concatenación de **R1** con **R2** y cuyas tuplas son todas las concatenaciones posibles de una tupla de **r1** con una tupla de **r2**.

Producto Cartesiano

- **Esquema R** = concatenación de **R1** y **R2**
- **Tuplas** = todas las tuplas **t** para las cuales existe **t1** en **r1**, y **t2** en **r2** tales que **t[R1] = t1[R1]** y **t[R2] = t2[R2]**

Se utiliza la siguiente notación para representar el producto cartesiano de dos relaciones **r1** y **r2**:

r1 X r2

Ejemplo de Producto cartesiano:

Dadas las relaciones **empleado(EMP)** y **departamento(DEPTO)**:

EMP	NROEMP	APELLIDO	NRODPTO	SUELDO
	4528	Gimenez	20	17500
	4356	Lopez	10	35890
	5678	García	10	8900

DEPTO	NRODPTO	NOMBRE
	10	Sistemas
	20	Finanzas
	30	Producción

fig.8

La relación resultante de la operación producto cartesiano **empleado X departamento** es:

NROEMP	APELLIDO	NRODTO	SUELDO	NRODTO	NOMBRE
4528	Gimenez	20	17500	10	Sistemas
4356	Lopez	10	35890	10	Sistemas
5678	García	10	8900	10	Sistemas
4528	Gimenez	20	17500	20	Finanzas
4356	Lopez	10	35890	20	Finanzas
5678	García	10	8900	20	Finanzas
4528	Gimenez	20	17500	30	Producción
4356	Lopez	10	35890	30	Producción
5678	García	10	8900	30	Producción

fig.9

Esquema **empleado X departameto** = (empleado.NROEMP,
empleado.APELLIDO,
empleado.NRODTO,
empleado.SUELDO,
departamento.NRODTO,
departamento.NOMBRE)

OPERACIONES ADICIONALES

Estas operaciones se deducen de las anteriores y son, por lo tanto, redundantes. No obstante, son esenciales para la creación práctica de los sistemas relacionales.

Por ejemplo, el producto cartesiano es una operación tan costosa en tiempo de ejecución y espacio físico requerido que es preferible realizar la operación de join.

Las operaciones adicionales que consideramos son las siguientes:

- * JOIN y sus casos particulares
 - a) equijoin
 - b) autojoin
 - c) natural join
- * INTERSECCION
- * DIVISION

JOIN

El join de dos relaciones **r1** y **r2** según un predicado **P** es una relación **r(R)** de esquema igual al del producto cartesiano **r1 X r2**, y cuyas tuplas son el conjunto de las tuplas de **r1 X r2** que satisfacen el predicado **P**.

En otros términos, podemos considerar al join de **r1** y **r2** según **P**, como la selección por un predicado **P** sobre **r1 X r2**.

Se empleará la siguiente notación para representar el join de dos relaciones **r1** y **r2** según **P**:

$$\boxed{r1 \bowtie_P r2}$$

Según lo expuesto:

$$r1 \bowtie_P r2 = \sigma_P(r1 \times r2)$$

Distinguiremos varios casos particulares del join:

- a) El **equi-join** de **r1** y **r2** sobre los atributos **r1.Ai** y **r2.Aj** es el join según el predicado **r1.Ai = r2.Aj**
- b) El **auto-join** de **r** sobre el atributo **Ai** es el join de **r** consigo misma según el predicado **1.Ai = 2.Ai**
- c) El **join natural** de **r1(R1)** y **r2(R2)** es una relación de esquema **R1 ∪ R2** con las tuplas del equi-join de **r1** y **r2** sobre todos los atributos que tienen el mismo nombre en **r1** y en **r2**.

Para indicar el join natural de **r1** y **r2** se utilizará la siguiente notación:

$$\boxed{r1 * r2}$$

Formalmente:

$$r1 * r2 = \pi_{R1 \cup R2}(r1 \bowtie_P r2)$$

Siendo **P** = (r1.A₁ = r2.A₁ ∧ r1.A₂ = r2.A₂ ∧ ... ∧ r1.A_n = r2.A_n)

y **R1 ∩ R2** = (A₁, A₂, ..., A_n)

Ejemplo de Join natural:

Dadas las relaciones **empleado(EMP)** y **departamento(DEPTO)** presentes en la figura 8, el join natural **empleado * departamento** tiene como relación resultante a:

NROEMP	APELLIDO	SUELDO	NRODPTO	NOMBRE
4356	Lopez	35890	10	Sistemas
5678	García	8900	10	Sistemas
4528	Gimenez	17500	20	Finanzas

fig.10

INTERSECCION

La intersección de dos relaciones **r1** y **r2** con el mismo esquema, es una relación **r** con igual esquema, que contiene las tuplas que pertenecen a ambas relaciones, es decir, a **r1** y **r2** a la vez.

Nótese que **r1** y **r2** deben ser *UNIONCOMPATIBLES*.

Para representar la intersección de dos relaciones **r1** y **r2** se utiliza la siguiente notación:

$$\boxed{r1 \cap r2}$$

La intersección entre dos relaciones se puede obtener a partir de la operación diferencia como:

$$r1 \cap r2 = r2 - (r2 - r1)$$

DIVISION

El cociente resultante de dividir la relación **r1**, de esquema **R1=(A₁,A₂,...,A_n)**, por la relación **r2**, de esquema **R2=(A_{p+1},...,A_n)**, es la relación **r**, de esquema **R=(A₁,A₂,A₃,...,A_p)**, formada por tuplas **t** tales que concatenada **t** con cada una de las tuplas de **r2** constituye una tupla existente en **r1**.

Nótese que, en términos de conjuntos, el esquema **R2** es un subconjunto de **R1**, y el esquema resultante **R** es la diferencia de conjuntos **R1 - R2**.

Si a_i es un valor cualquiera del atributo A_i podemos definir el cociente r , resultante de dividir $r1$ por $r2$ como:

$$r = \{(a_1, a_2, \dots, a_p) / \forall (a_{p+1}, \dots, a_n) \in r2 : \exists (a_1, a_2, \dots, a_p, a_{p+1}, \dots, a_n) \in r1\}$$

Para representar la división de dos relaciones $r1$ y $r2$ se utiliza la siguiente notación:

$$r1 \% r2$$

La división se puede obtener a partir de la diferencia, el producto cartesiano y la proyección de la forma siguiente:

$$r1 \% r2 = u1 - u2$$

siendo

$$u1 = \pi_{A1, A2, \dots, AP}(r1)$$

$$u2 = \pi_{A1, A2, \dots, AP}((u1 \times r2) - r1)$$

Ejemplo de División de relaciones:

Sean las relaciones $r1(R1)$, $r2(R2)$ y $r3(R3)$:

R1	S#	P#	Q#
	S1	P1	Q1
	S1	P2	Q2
	S1	P3	Q3
	S2	P1	Q1
	S2	P2	Q1
	S2	P3	Q3
	S3	P3	Q2
	S3	P4	Q1
	S4	P1	Q2
	S4	P2	Q2

R2	P#	Q#
	P1	Q1
	P2	Q2
	P3	Q3

R3	P#
	P1
	P2

fig.11

La figura 12 muestra los respectivos resultados de las operaciones $r1 \% r2$ y $r1 \% r3$:

$r1 \% r2$	
S#	
S1	

$r1 \% r3$	
S#	Q#
S2	Q1
S4	Q2

fig.12

COMPOSICION DE RELACIONES

Las operaciones presentadas, encadenadas sucesivamente y aplicadas a las relaciones, permiten elaborar consultas complejas sobre una base de datos relacional.

A continuación presentaremos un conjunto de ejemplos que permiten visualizar la construcción de expresiones necesarias para consultar una base de datos relacional.

Sean las relaciones:

pedido (PEDIDO)
producto (PRODUCTO)
proveedor (PROVEEDOR)

PROVEEDOR	PROVE#	APELLIDO
	128	Gomez
	547	García
	328	Valenti

PRODUCTO	P#	NOMBRE	COLOR
	12	Rueda	Rojo
	33	Tuerca	Azul
	22	Perno	Rojo
	34	Tuerca	Verde

PEDIDO	P#	PROVE#	CANTIDAD
	34	547	100
	33	547	150
	12	128	360
	34	328	230
	12	547	70
	22	328	200
	22	547	240

fig.13

- a) Obtener los nombres de aquellos proveedores que suministran el producto número 34.
 Planteamos la solución paso a paso:

$$\begin{aligned} r1 &= \pi_{P\#, APELLIDO} (\text{proveedor} * \text{pedido}) \\ r2 &= \sigma_{P\# = 34}(r1) \\ r &= \pi_{APELLIDO} (r2) \end{aligned}$$

Usando una expresión compuesta:

$$r = \pi_{APELLIDO} (\sigma_{P\# = 34}(\pi_{P\#, APELLIDO} (\text{proveedor} * \text{pedido})))$$

- b) Obtener los números de proveedor de aquellos proveedores que suministran al menos un producto de color rojo.
 Solución:

$$r = \pi_{PROVE\#} (\sigma_{COLOR = "Rojo"} (\text{producto} * \text{pedido}))$$

- c) Obtener los nombres de los proveedores que suministran todos los productos.
 Solución:

$$\begin{aligned} r1 &= \pi_{PROVE\#, P\#} (\text{pedido}) \% \pi_{P\#} (\text{producto}) \\ r2 &= r1 * \text{proveedor} \\ r &= \pi_{APELLIDO} (r2) \end{aligned}$$

APENDICE

TRANSFORMACION DEL MODELO E-R A OTROS MODELOS IMPLEMENTABLES

El problema que se nos presenta es el siguiente: dado un Modelo Conceptual con la forma de un MER ¿cómo diseñar estructuras para representar los datos correspondientes?.

En el Modelo Conceptual, las interrelaciones son pares (o ternas, etc.) abstractos, que relacionan puntos abstractos de conjuntos. En los Modelos Operacionales es preciso transformar esas abstracciones en datos concretos, representables en un computador.

En general, cada conjunto de entidades E estará representado en el Modelo Operacional por un conjunto de datos D . Cada entidad de E estará representada por un elemento de D .

En lo que refiere a los conjuntos de interrelaciones, existen dos maneras, según los distintos Modelos Operacionales, de representar las interrelaciones conceptuales:

- la representación *Implícita*: sin acrecentar los datos
- la representación *Explícita*: introduciendo datos específicos para representar la interrelación

TRANSFORMACION DE UN MODELO E-R A TABLAS

TIPOS DE ENTIDADES REGULARES

Sea E un conjunto de entidades fuertes con atributos descriptivos A_1, A_2, \dots, A_n . Podemos representar este conjunto de entidades mediante una *tabla* T con n *columnas*, cada una de las cuales corresponde a un atributo de E . Cada *fila* (tupla) en esta tabla corresponde a una entidad del conjunto de entidades E .

Ejemplo:

PAÍS (nombre_país, cant_hab_país)

TALLER (nro_taller, dirección)

Países

NOMBRE_PAÍS	CANT_HAB_PAÍS

talleres

NRO_TALLER	DIRECCIÓN

TIPOS DE ENTIDADES DÉBILES

Sea **D** un conjunto de entidades débiles con atributos descriptivos D_1, D_2, \dots, D_n . Sea **E** el conjunto de entidades fuertes del que depende **D**. Sean A_1, A_2, \dots, A_k los atributos que constituyen la clave primaria de **E**. Podemos representar el conjunto de entidades **D** mediante una *tabla T* con $n + k$ columnas, cada una de las cuales corresponde a un atributo del conjunto $\{(A_1, A_2, \dots, A_k) \cup (D_1, D_2, \dots, D_n)\}$.

Ejemplo:

PAÍS (nombre_país, cant_hab_país)

PROVINCIA (nombre_provincia, cant_hab_provincia)

Provincias

NOMBRE_PAÍS	NOMBRE_PROVINCIA	CANT_HAB_PROVINCIA

TIPOS DE INTERRELACION

Sea **R** un conjunto de interrelaciones n-arios entre los conjuntos de entidades E_1, E_2, \dots, E_n .

Sea $\{K E_i\}$ el conjunto de atributos que conforman la clave primaria del conjunto de entidades E_i .

Sean A_1, A_2, \dots, A_p los atributos descriptivos propios de la interrelación **R**. Podemos representar el conjunto de interrelaciones **R** mediante una *tabla T* con una *columna* por cada atributo del siguiente conjunto :

$$\left\{ \bigcup_{i=1}^n \{K E_i\} \cup (A_1, A_2, \dots, A_p) \right\}$$

Ejemplo:

TREN (nro_tren, denominación)

FECHA (día, mes, año)

TRAMO (nro_tramo, distancia)

VIAJE_TRAMO (cant_asientos_disponibles, horario_llegada, horario_salida)

Viaje_tramo

NRO_TREN	NRO_TRAMO	DIA	MES	AÑO	CANT_ASIENT_DISP	H_LLEGA	H_SALE

GENERALIZACION Y ESPECIALIZACION

Existen dos variantes para la Generalización, sólo la primera de ellas es válida también para la Especialización :

1. Crear una *tabla* para el tipo de entidad de mayor nivel (más general), tal como se especificó para los tipos de entidades regulares. Crear una *tabla* para cada tipo de entidad de menor nivel (menos general), que incluya una *columna* para cada uno de los atributos descriptivos, más una *columna* por cada atributo de la clave primaria del tipo de entidad más general.

Ejemplo:

EMPLEADO (nro_empleado, nombre_empleado, dirección_empleado)

EMPLEADO_TALLER (especialidad)

EMPLEADO_TREN (jornada, tipo_trabajo)

DIRECTIVO (profesión, cargo)

Empleados

NRO_EMPLEADO	NOMBRE_EMPLEADO	DIRECCION_EMPLEADO

Empleados_taller

NRO_EMPLEADO	ESPECIALIDAD

Empleados_tren

NRO_EMPLEADO	JORNADA	TIPO_TRABAJO

Directivos

NRO_EMPLEADO	PROFESION	CARGO

2. Crear una tabla para cada tipo de entidad de menor nivel (menos general), que incluya una columna para cada uno de sus atributos descriptivos, más una columna por cada atributo del tipo de entidad más general. En este caso no se crea tabla para el tipo de entidad de mayor nivel (más general).

Ejemplo:EMPLEADO (nro_empleado, nombre_empleado, dirección_empleado)

EMPLEADO_TALLER (especialidad)

EMPLEADO_TREN (jornada, tipo_trabajo)

DIRECTIVO (profesión, cargo)

Empleados_taller

NRO_EMPLEADO	NOMBRE_EMPLEADO	DIR_EMPLEADO	ESPECIALIDAD

Empleados_tren

NRO_EMPLEADO	NOMBRE_EMPLEADO	DIR_EMPLEADO	JORNADA	TIPO_TRABAJO

Directivos

NRO_EMPLEADO	NOMBRE_EMPLEADO	DIR_EMPLEADO	PROFESION	CARGO