



Metodología para el Modelado de Datos

Base de Datos

Introducción

- ▶ En clases anteriores se han visto los principios de modelado a través del uso de **abstracciones**
- ▶ También se ha visto el **Modelo Entidad Relación** como una forma idónea de representación de modelos de datos
- ▶ El proceso de creación de un modelo se puede ver como un proceso de tipo incremental, se parte de una representación que es refinada y enriquecida gradualmente
- ▶ Veremos determinadas primitivas de refinamiento, un conjunto de transformaciones que dado un esquema inicial proponen uno final

Metodologías

- ▶ Para ello aplicaremos estrategias de diseño:
 - ▶ **Descendentes**, se parte de conceptos generales abstractos y se refinan para hacerlos concretos
 - ▶ **Ascendentes**, se parte de conceptos concretos y se abstraen
 - ▶ **Mixtas**, para cada concepto se elige entre las anteriores
 - ▶ **Centrifugas**, se comienza con el concepto central y se continua con conceptos relacionados
- ▶ Toda metodología debe tener dos componentes: ser **rigurosa** y ser **flexible**

Metodologías de Diseño

- ▶ El **Diseño** es el resultado de un análisis complejo de los requerimientos de los usuarios
- ▶ Se produce a través de un proceso iterativo, comenzando con un **esquema inicial** y, mediante un **conjunto de transformaciones**, se obtiene un **esquema definitivo**
- ▶ Las **transformaciones** tienen tres características básicas:
 - ▶ Toda transformación tiene un esquema inicial y un esquema final resultante de la transformación
 - ▶ Toda transformación crea una relación entre nombre de conceptos iniciales y nombres de conceptos resultantes
 - ▶ El concepto resultante debe heredar todas las conexiones lógicas definidas en el esquema inicial

Ejemplo...

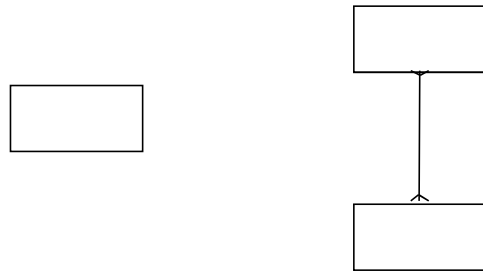
- ▶ Se quiere guardar información de la localidad donde ha nacido cada uno de los empleados de la empresa, teniendo en cuenta que:
 - ▶ Un empleado ha nacido en una sola localidad.
 - ▶ Cada localidad pertenece a una única provincia.
 - ▶ Cada provincia pertenece a una única región del país.

Transformaciones primitivas

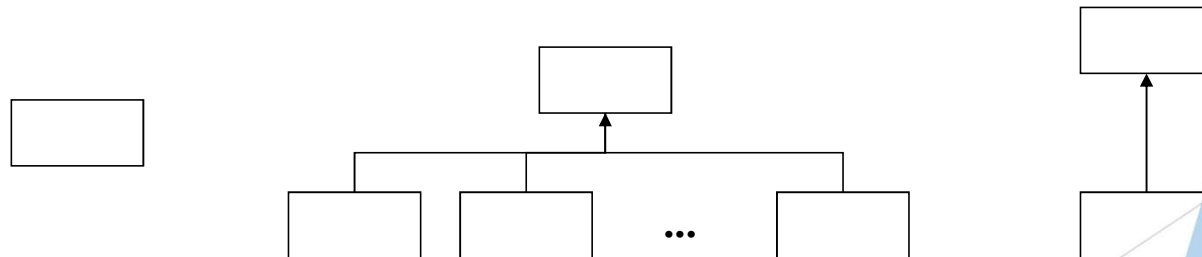
- ▶ La idea básica es conseguir un conjunto pequeño de transformaciones que puedan ser usadas para convertir, con eficiencia, esquemas iniciales en esquemas finales
- ▶ Se propone un conjunto de **transformaciones primitivas**, con una estructura simple, obteniendo conceptos que no se pueden transformar en otros más simples. Las transformaciones primitivas se pueden clasificar en **descendientes** y **ascendentes**
 - ▶ **Descendientes**, son refinamientos puros, dado un concepto simple se obtiene una descripción más detallada del concepto
 - ▶ **Ascendentes**, producen conceptos y propiedades nuevas que no aparecían en versiones anteriores

Primitivas Descendentes

1. Entidad -> Entidades relacionadas



2. Entidad -> Generalización (Entidad-> Subconjunto)

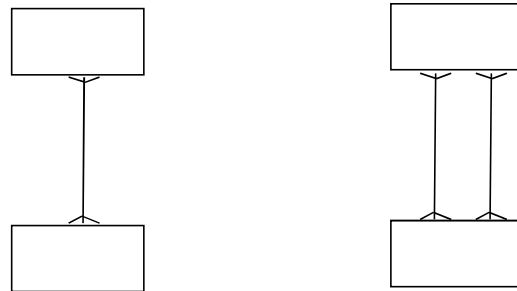


Primitivas Descendentes

3. Entidad -> Entidades no relacionadas

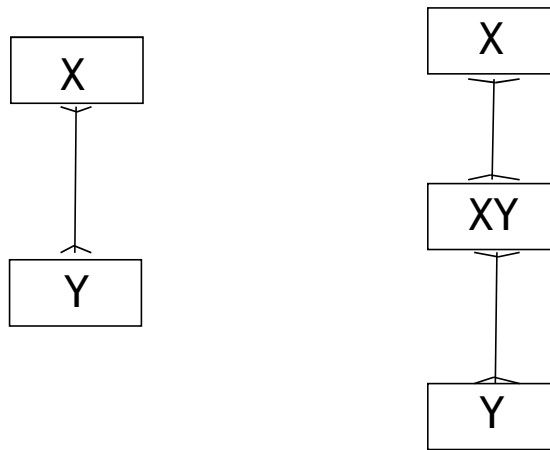


4. Relación -> Relaciones Paralelas



Primitivas Descendentes

5. Relación -> Entidad con relaciones

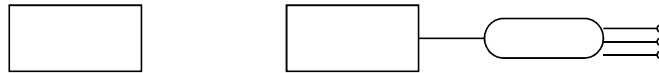


6. Desarrollo de Atributos



Primitivas Descendentes

7. Desarrollo de Atributos Compuestos



8. Refinamiento de Atributos

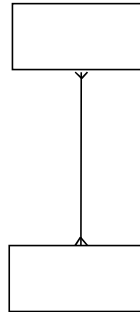


Primitivas Ascendentes

1. Generación de Entidad

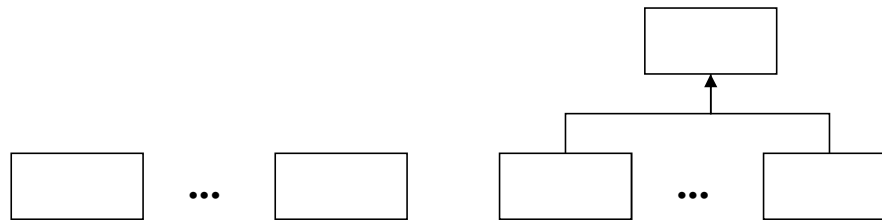


2. Generación de Relación



Primitivas Ascendentes

3. Generación de Generalización



4. Agregación de Atributos



5. Agregación de Atributos Compuestos



Propiedades de las Primitivas

- ▶ **Compleción**, un conjunto de primitivas es completo si cualquier esquema de base de datos se puede construir a partir de uno vacío, mediante la aplicación en secuencia de las primitivas
- ▶ **Minimalidad**, un conjunto de primitivas es mínimo si ninguna de ellas se puede expresar usando las otras primitivas
- ▶ Nuestras primitivas:
 - ▶ ¿las primitivas descendentes son mínimas?
 - ▶ ¿las primitivas descendentes son completas?
 - ▶ ¿las primitivas ascendentes son mínimas?
 - ▶ ¿las primitivas ascendentes son completas?

Aplicación

BASE de DATOS DEMOGRÁFICA

- ▶ En la base de datos de un censo se consideran las siguientes propiedades de las personas: nombre, apellido, sexo, edad, lugar de nacimiento, lugar de residencia, años de residencia, situación militar de los hombres, apellido de soltera de las mujeres.
- ▶ Los lugares pueden ser estados extranjeros o ciudades nacionales. Cada uno tiene un nombre y número de habitantes (que representa la población total en el caso de los estados extranjeros) y los nombres de las regiones o ciudades.

Comparación de Metodologías

► **Descendente**

- Los Conceptos se refinan progresivamente
- **Ventajas**
 - No hay efectos secundarios indeseables
- **Desventajas**
 - Requiere un diseñador hábil con alta capacidad de abstracción desde el comienzo

► **Ascendente**

- Los conceptos se crean a partir de componentes elementales
- Decisiones locales de diseño sencillas
- **Ventajas**
 - Ninguna carga para el diseñador inicial
- **Desventajas**
 - Necesidad de una reestructuración después de aplicar las primitivas ascendentes

Comparación de Metodologías

▶ **Centrífuga**

- ▶ Los conceptos se construyen con un enfoque de “**mancha de aceite**”
- ▶ **Ventaja**
 - ▶ Facilidad para descubrir nuevos conceptos cercanos a los anteriores
 - ▶ Ninguna carga sobre el diseñador inicial
- ▶ **Desventaja**
 - ▶ La visión global del dominio de aplicación se obtiene sólo al final

▶ **Mixta**

- ▶ Partición descendente de los requerimientos, integración ascendente usando el esquema de armazón
- ▶ **Ventaja**
 - ▶ Enfoque de “**divide y vencerás**”
- ▶ **Desventajas**
 - ▶ Requiere decisiones cruciales sobre el esquema armazón al inicio del proceso de diseño

Metodología para el modelado de datos

- ▶ Cuando el problema está definido en lenguaje natural, es necesario aplicar un conjunto de pasos de diseño para obtener éxito en el modelado:
 - ▶ **Análisis de Requerimientos**
 - ▶ **Diseño inicial**, construyendo un esquema almacén
 - ▶ **Diseño completo**, para cada concepto del esquema almacén, aplicar:
 - ▶ **Primitivas descendentes**
 - ▶ **Primitivas ascendentes**
 - ▶ Hasta que todos los conceptos queden expresados en el esquema final

Análisis de Requerimientos

- ▶ Debemos analizar cuidadosamente el texto proporcionado para detectar **ambigüedades** y entender el significado de los términos detalladamente:
 - ▶ Elegir el nivel apropiado de abstracción de los términos
 - ▶ Evitar el uso de casos particulares en lugar de conceptos generales
 - ▶ Evitar las expresiones vagas o indirectas
 - ▶ Elegir un estilo estándar para los enunciados
 - ▶ Verificar los homónimos y sinónimos
 - ▶ Hacer explícitas las referencias entre términos
 - ▶ Utilizar un glosario

Análisis de requerimientos

1 En una base de datos de una universidad, se representan datos
2 sobre estudiantes y profesores. Para los estudiantes, se
3 representa el apellido, edad, sexo, ciudad y provincia de
4 nacimiento, ciudad y provincia de residencia de sus
5 familias, lugares y provincias donde vivieron antes
6 (con el lapso que vivieron en cada uno), cursos que han
7 aprobado, con nombre, código, profesor,
8 nota y fecha. Asimismo, se representan los cursos
9 a los que asisten en la actualidad y, para cada uno, día, sitios
10 y horas de impartición de las clases (cada curso
11 se imparte a lo sumo una vez en un día). Para estudiantes graduados,
12 se representa el nombre del consejero
13 y el número total de créditos en el último año.
14 Para estudiantes de doctorado, se representa el título y área
15 de investigación de su tesis. Para los maestros, se
16 representa el apellido, edad, lugar y provincia de nacimiento,
17 nombre del departamento al que pertenecen, número de teléfono,
18 título, situación y temas de investigación.

Análisis de requerimientos

1 En una base de datos de una universidad, se representan datos
2 sobre estudiantes y profesores. Para los estudiantes, se
3 representa el apellido, edad, sexo, ciudad y provincia de
4 nacimiento, ciudad y provincia de residencia de sus
5 familias, lugares y provincias donde vivieron antes
6 (con el lapso que vivieron en cada uno), cursos que han
7 aprobado, con nombre, código, profesor,
8 nota y fecha. Asimismo, se representan los cursos
9 a los que asisten en la actualidad y, para cada uno, día, sitios
10 y horas de impartición de las clases (cada curso
11 se imparte a lo sumo una vez en un día). Para estudiantes graduados,
12 se representa el nombre del consejero
13 y el número total de créditos en el último año.
14 Para estudiantes de doctorado, se representa el título y área
15 de investigación de su tesis. Para los maestros, se
16 representa el apellido, edad, lugar y provincia de nacimiento,
17 nombre del departamento al que pertenecen, número de teléfono,
18 título, situación y temas de investigación.

- ▶ Elegir el nivel apropiado de abstracción de los términos
- ▶ Evitar el uso de casos particulares en lugar de conceptos generales
- ▶ Evitar las expresiones vagas o indirectas
- ▶ Elegir un estilo estándar para los enunciados
- ▶ Verificar los homónimos y sinónimos
- ▶ Hacer explícitas las referencias entre términos
- ▶ Utilizar un glosario

Análisis de requerimientos

► Nivel apropiado de abstracción

- 1 En una base de datos de una universidad, se representan datos
- 2 sobre estudiantes y profesores. Para los estudiantes, se
- 3 representa el apellido, edad, sexo, ciudad y provincia de
- 4 nacimiento, ciudad y provincia de residencia de sus
- 5 familias, lugares y provincias donde vivieron antes
- 6 (con el **lapso** que vivieron en cada uno), cursos que han
- 7 aprobado, con nombre, código, profesor,
- 8 nota y fecha. Asimismo, se representan los cursos
- 9 a los que asisten en la actualidad y, para cada uno, día, sitios
- 10 y horas de impartición de las clases (cada curso
- 11 se imparte a lo sumo una vez en un día). Para estudiantes graduados,
- 12 se representa el nombre del consejero
- 13 y el número total de créditos en el último año.
- 14 Para estudiantes de doctorado, se representa el título y área
- 15 de investigación de su tesis. Para los maestros, se
- 16 representa el apellido, edad, **lugar** y provincia de nacimiento,
- 17 nombre del departamento al que pertenecen, número de teléfono,
- 18 título, **situación** y temas de investigación.

Análisis de requerimientos

- ▶ Uso de casos particulares en lugar de conceptos generales
 - ▶ Modelando un área de Helpdesk hago referencia a teclados, monitores, etc en lugar de “Partes o Componentes de PC”
- ▶ Evitar expresiones vagas o indirectas
 - ▶ Persona sentada en la recepción del edificio - Portero
- ▶ Elegir estilo estandarizado de enunciados (Pero expresivo ojo)
 - ▶ <sujeito> <verbo> <especificación>
 - ▶ <si> <condición> <entonces> <acción>
- ▶ Verificar sinónimos y homónimos

●Base de Datos - Metodología para el Modelado de Datos

Homónimos

Palabras que se pronuncian o escriben igual, pero tienen distintos significados

 ¡Qué hermoso es tu **traje**!
Lo **traje** todo en la maleta.

 Su **cara** reflejaba felicidad.
María no compra ropa **cara**.

Análisis de requerimientos

- ▶ Hacer explícitas las referencias entre términos
 - ▶ Día es fecha?
- ▶ Creación de glosario
 - ▶ Buena forma de eliminar ambigüedades
 - ▶ Lleva tiempo y trabajo
 - ▶ Deberían usarse luego esos términos solamente

Análisis de requerimientos

Línea	Término	Nuevo término	Razones para la corrección
5	Lugares	Ciudades	<i>Lugar</i> es una palabra genérica
6	Lapso	Núm. de años	<i>Lapso</i> es una palabra genérica
9	Actualidad	Año actual	<i>Actualidad</i> es ambigua
9	Día	Día de la semana	Más específico
9	Sitios	Aulas	Homónimo de <i>lugares</i> en la línea 5
10	Clases	Cursos	Sinónimo de <i>cursos</i> en la línea 8
15	Maestro	Profesor	Sinónimo de <i>profesor</i> en la línea 2
16	Lugar	Ciudad	Lo mismo que en la línea 5
17	Teléfono	Teléfono del departamento	Más específico
18	Situación	Estado civil	<i>Situación</i> es ambiguo
18	Tema	Area de investigación	Sinónimo de <i>área de investigación</i> en la línea 15

Análisis de requerimientos

En una base de datos de una universidad se representan datos sobre estudiantes y profesores.

Enunciados generales

Para los estudiantes, se representa el apellido, edad, sexo, ciudad y provincia de nacimiento, ciudad y provincia de residencia de sus familias, ciudades y provincias donde han vivido antes (con el número de años que vivieron en cada una), cursos que han aprobado, con nombre, código, profesor, nota y fecha.

Enunciados sobre estudiantes

Asimismo se representan los cursos a los que asisten en el año actual y, para cada uno, día de la semana, aulas y horas de impartición de los cursos (cada curso se imparte a lo sumo una vez en un día).

Enunciados sobre cursos

Para estudiantes graduados se representa el nombre del consejero en el último año. Para estudiantes de doctorado se representa el título y área de investigación de sus tesis.

Enunciados sobre tipos específicos de estudiantes

Para profesores se representa el apellido, edad, ciudad y provincia de nacimiento, nombre del departamento al que pertenecen, número de teléfono del departamento, título, estado civil y área de investigación.

Enunciados sobre profesores

Análisis de Requerimientos

- ▶ A partir del enunciado, elaborar dos listas:
 - ▶ Candidatos a entidades
 - ▶ Candidatos a relaciones (si es posible identificar cardinalidades, mejor!)
- ▶ Construir una **Matriz de Entidades y Relaciones** como la siguiente:
- ▶ Pasar a un diagrama de MER aproximado

	E_1	E_2	E_3	...	E_n
E_1	REL ₁	REL ₂		...	REL ₃
E_2		REL ₄	REL ₅	...	
E_3			REL ₆	...	
...				...	
E_n				...	REL _n

Ejercicio

Se trata de una base de datos para un entorno de programación. En este entorno los programadores producen programas, que se escriben en determinados lenguajes de programación. Cada programa es escrito por un determinado programador, y puede llamar a otros programas y puede ser utilizado por determinados usuarios. Los usuarios se reconocen por su nombre de entrada al sistema y una contraseña. Los programas incluyen el nombre del programa, la extensión y el código del programador. Los programas tienen un número de versión, una fecha y una breve descripción; algunos programas interactúan con la base de datos. Cada DBMS mantiene datos almacenados en forma de relaciones, con varios atributos y una clave primaria. Cada base de datos la define un DBA que es un programador especializado en su administración.

Ejercicio - Enunciados

Se trata de una base de datos para un entorno de programación. En este entorno los programadores producen programas, que se escriben en determinados lenguajes de programación.

Enunciados Generales

Cada programa es escrito por un determinado programador, y puede llamar a otros programas y puede ser utilizado por determinados usuarios. Los programas incluyen el nombre del programa, la extensión y el código del programador. Los programas tienen un número de versión, una fecha y una breve descripción; algunos programas interactúan con la base de datos.

Enunciados de Programas

Los usuarios se reconocen por su nombre de entrada al sistema y una contraseña.

Enunciados de usuarios

Cada DBMS mantiene datos almacenados en forma de relaciones, con varios atributos y una clave primaria. Cada base de datos la define un DBA que es un programador especializado en su administración.

•Base de Datos - Metodología para el Modelado de Datos

Enunciados DBMS

Ejercicio - Matriz de ER

	PROGRAMAS	PROGRAMADORES	USUARIOS	DBMS
PROGRAMAS	REL ₁	REL ₂		...
PROGRAMADORES		REL ₄	REL ₅	...
USUARIOS			REL ₆	...
DBMS				...

Conflictos en la Integración de Modelos

- ▶ Dependiendo de la metodología utilizada puede ocurrir que haya que integrar dos o más modelos en uno, pudiendo ocurrir conflictos, causados por:
 - ▶ **Perspectivas diferentes**, cada uno de los modelos a integrar se modelaron utilizando puntos de vista distintos
 - ▶ **Equivalencias entre las construcciones del modelo**, puede ser que hayan distintas representaciones de la misma realidad, causado por la rica variedad de estructuras de representación
 - ▶ **Especificaciones de diseño incompatibles**, puede que los modelos contengan errores de nombre, estructuras y restricciones de integridad, debiéndose detectar y corregir en la etapa de análisis y resolución de conflictos.

Análisis y Resolución de Conflictos

- ▶ Se intenta detectar todas las diferencias de representación en los modelos involucrados:
 - ▶ **Análisis de Conflictos de Nombre**, comparando y unificando nombres, se basa en semejanzas y discrepancia de conceptos:
 - ▶ **Semejanzas de Propiedades**, es decir: atributos, relaciones, subconjuntos, generalizaciones, comunes o similares
 - ▶ **Semejanza de Restricciones**, o sea reglas comunes o similares que limitan el conjunto de casos válidos del esquema
 - ▶ Por lo tanto tenemos:
 - ▶ **Sinónimos**, para los casos de semejanza
 - ▶ **Homónimos**, para los casos de discrepancia

Análisis y Resolución de Conflictos

- ▶ **Análisis de Conflictos Estructurales**, se buscan partes de los esquemas con los mismos nombre y se tratan de fusionar:
 - ▶ **Conceptos idénticos**, se unifican
 - ▶ **Conceptos compatibles**, son estructuras o propiedades no contradictorias, se altera una de las dos representaciones
 - ▶ **Conflictos incompatibles**
 - ▶ **Cardinalidades diferentes**, para los atributos o entidades
 - ▶ **Identificadores diferentes**, para entidades y relaciones
 - ▶ **Subconjuntos invertidos**

Transformación de Esquemas

- ▶ Podemos clasificar las transformaciones en:
 - ▶ **Transformaciones que preservan la información**, como son las transformaciones para aumentar la legibilidad
 - ▶ **Transformaciones que cambian la información**
 - ▶ **Transformaciones de aumento**, como son el diseño descendente o ascendente
 - ▶ **Transformaciones de reducción**, eliminado conceptos superfluos que no están en los requerimientos y deben ser eliminados
 - ▶ **Transformaciones no comparables**, como son las causadas por la resolución de conflictos

Transformaciones para lograr Minimalidad

- ▶ Suele ocurrir que los requerimientos de los problemas contienen redundancia en la definición y se suele trasladar a los esquemas
- ▶ En los metodologías mixtas también se suelen expresar redundancias que se deben resolver en la integración
- ▶ **Ciclos de Relaciones**, se da cuando una relación tiene le mismo contenido de información que otro conjunto de relaciones
 - ▶ La redundancia depende del significado de los conceptos, puede que dos caminos tengan significado distinto o quizás igual
 - ▶ Se pueden hacer comprobaciones con respecto a la cardinalidad para detectar las redundancia
 - ▶ Los ciclos también puede representar algún tipo de restricción de “contenida en”.

Transformaciones para lograr Minimalidad

- ▶ ***Atributos derivados***, atributos que se deducen por algún algoritmo
- ▶ ***Subconjuntos implícitos***, subconjuntos contenidos en otros.

Transformaciones para lograr expresividad

- ▶ **Eliminación de subentidades colgantes** en las jerarquías de generalización, si luego de obtener el modelo los subconjuntos resultantes no tienen propiedades se puede sustituir por un atributo en la entidad generalizada. Se debe pesar entre la expresividad y la autoexplicación
- ▶ **Eliminación de entidades colgantes**, son aquellas que tienen pocos atributos y se relacionan con una entidad más importante, se podría eliminar la entidad y la relación y pasar los atributos a la entidad
- ▶ **Creación de una generalización**, se aplica cuando se descubre dos entidades con propiedades similares, que pertenecen a la misma jerarquía de generalización
- ▶ **Creación de un subconjunto**, destaca el papel de una entidad, se debe realizar cuando el subconjunto tiene una identidad clara y significativa para el diseño

Transformaciones para lograr normalidad

- ▶ La **normalidad** se puede utilizar para verificar un correcto diseño del esquema
- ▶ El Modelo Entidad Relación tiende a producir esquemas normalizados:
 - ▶ El objetivo de la normalización es mantener las DF separadas, al asociar con cada grupo de DF homogéneas un elemento del modelo, que tenga como identificador la DF
 - ▶ Si cada concepto del dominio de aplicación corresponde con exactamente un concepto del esquema, que será identificado por su DF, entonces el esquema estará normalizado

Validación de los Diseños

- ▶ Para dar por finalizado el proceso de modelado se debe validar, revisando las cualidades del esquema:
- ▶ **Compleción:** un esquema es completo cuando representa todas las características del dominio de aplicación
 - ▶ Se puede verificar mirando la especificación de requerimientos y verificando que está en algún lugar del esquema
 - ▶ Se puede revisar el esquema y verificar que cada concepto se menciona en la especificación
 - ▶ También se puede verificar contra el diseño funcional del problema

Validación de los Diseños

- ▶ **Corrección:** un esquema es correcto cuando usa adecuadamente los conceptos del MER
 - ▶ **Corrección sintáctica:** cuando los conceptos se definen con propiedad en el modelo
 - ▶ **Corrección semántica:** cuando los conceptos se utilizan acorde a su definición
 - ▶ Errores más comunes:
 - ▶ Usar un atributo en lugar de una entidad
 - ▶ Omitir una generalización
 - ▶ Olvidar la propiedad de herencia de las generalizaciones
 - ▶ Usar una relación con número erróneo de entidades
 - ▶ Usar una entidad en lugar de una relación
 - ▶ Olvidar identificadores
 - ▶ Omitir la cardinalidad

Validación de los Diseños

- ▶ **Minimalidad**, un esquema es mínimo cuando cada aspecto de los requerimientos aparece una sola vez en el esquema:
 - ▶ No se puede borrar ningún concepto del esquema sin perder alguna información
 - ▶ En el caso que se permita mantener alguna redundancia en el esquema, ésta se debe documentar
- ▶ **Expresividad**, un esquema es expresivo cuando la representación de los conceptos es de forma natural y se pueden entender sin necesidad de explicaciones adicionales.

Validación de los Diseños

- ▶ **Legibilidad**, un diagrama tiene buena legibilidad cuando respeta determinados aspectos estéticos
 - ▶ Tamaño aproximadamente igual de las entidades y relaciones
 - ▶ Destacar estructuras simétricas
 - ▶ Minimizar el número de cruces
 - ▶ Minimizar los quiebre de líneas
 - ▶ Situar los padres encima de los hijos en las generalizaciones
- ▶ **Autoexplicación**, un esquema es autoexplicativo cuando las propiedades representadas se reconocen sin la necesidad de anotaciones adicionales

Validación de los Diseños

- ▶ **Extensibilidad**, un modelo se adapta fácilmente a nuevos requerimientos si se puede descomponer en partes, y aplicar los cambios a las mismas.
- ▶ **Normalidad**, pretende que el esquema cumpla con las formas normales
 - ▶ Representaciones de cardinalidades mínimas y máximas
 - ▶ Representaciones de identificadores

Metodologías para el modelado de datos

¿Dudas?

