Universidad Central "Marta Abreu" de Las Villas

Facultad de Matemática, Física y Computación

Métodos de validación estructural de esquemas conceptuales Entidad-Relación

Tesis en opción al título de Ingeniero Informático

Autora: Lisbet Aguilar Guerra

Tutor: Dr. C. Carlos E. García González

Santa Clara, 2011

Hago constar que el presente trabajo fue realizado en la Universidad Central "Marta Abreu de Las Villas" como parte de la culminación de los estudios de la especialidad de Ciencias de la Computación, autorizando a que el mismo sea utilizado por la institución, para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos ni publicado sin la autorización de la Universidad.

Firma del autor

Los abajo firmantes, certificamos que el presente trabajo ha sido realizado según acuerdos de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

Firma del tutor

Firma del jefe del Seminario

AGRADECIMIENTOS

A Carlos García por haber sido mi profesor y ahora tutor de la tesis de grado, quien me ha ayudado y apoyado en todo momento.

A mis padres por haber hecho de mí la persona que soy.

A mis hermanos por quererme y apoyarme siempre incondicionalmente.

A Manuel por estar a mi lado a lo largo del camino y brindarme su amor y comprensión.

A mi familia y la de Manuel por apoyarme y ser tan especiales.

A todos mis amigos, compañeros de aula y profesores que de alguna manera han contribuido con mi educación profesional.

DEDICATORIA

A mi mamá y a mi papá que me han dedicado lo más importante y maravilloso que tiene la vida: Amor.

SÍNTESIS

Durante el procedo de creación del esquema conceptual de una base de datos, el diseñador puede incurrir en errores e inconsistencias que pueden propagarse a las siguientes fases del diseño, por lo que resulta importante identificarlas en las fases iniciales del diseño. Por otra parte la mayoría de las herramientas de diseño de bases de datos no realizan la validación del esquema conceptual, a pesar de que en la literatura se reportan trabajos relacionados con métodos de validación de esquemas conceptuales.

En la presente tesis se realiza un análisis de dos de estos métodos y se conformaron casos de estudio a los cuales se les aplicaron dichos métodos para comprobar su eficacia y determinar cuál abarca mayor cantidad de tipos de construcciones con el objetivo de recomendar su inclusión en una herramienta CASE.

ABSTRACT

During the procedure of creating the conceptual schema of a database, the designer can make mistakes and inconsistencies that can spread to the next phases of design, so it is important to identify in the early stages of design. Moreover most of the design tools databases do not perform validation of the conceptual scheme, although reported in the literature work on methods for validation of conceptual schemes.

In this thesis, an analysis of two of these methods and case studies were content to which we applied these methods to verify its effectiveness and determine which covers more types of buildings in order to recommend its inclusion in a CASE tool.

TABLA DE CONTENIDOS

INTRODUCCIÓN

La primera aproximación para describir la estructura de los datos para ser manipulada por un Sistema de Gestión de Bases de Datos (SGBD) se conduce con la ayuda de algún modelo conceptual, cuyos conceptos, valga la redundancia, deben ser naturales o al menos comprensibles para los usuarios del universo de discurso; término usado con mucha frecuencia para delimitar un área del mundo real que es objeto de modelación.

Estos modelos deben ofrecer una rica y variada colección de conceptos de modelación con énfasis en las interrelaciones lógicas entre los datos. El objetivo de tal modelación es encontrar los conceptos esenciales presentes en la fase de análisis de requerimientos. Esta tarea no es trivial pues a veces los usuarios describen un mismo concepto con términos diferentes (sinónimos) o usan un mismo término para conceptos diferentes (homónimos) º(García and Montes de Oca 2005). El resultado del diseño conceptual es el *esquema conceptual* de la base de datos, el cual es una descripción de alto nivel de la estructura de la base de datos, y no es más que un conjunto estático de representaciones lingüísticas y gráficas, invariables en el tiempo, que describen la estructura de los datos (Batini, Ceri et al. 1992).

El modelo conceptual de datos más popular, sin lugar a dudas, sigue siendo el modelo Entidad Relación (ER, acrónimo del inglés <u>Entity Relationship</u>)[1](Chen 1976) . Los conceptos que le sirven de base a este modelo son los de entidad e interrelación. La decisión de si un concepto del universo de discurso debe ser una entidad, una interrelación, o simplemente una propiedad de cualquiera de ellos, no siempre es obvia y en efecto puede existir más de una descripción conceptual que refleje los requerimientos de los usuarios. Se reconoce que la experiencia del diseñador es probablemente la mejor ayuda para tomar la decisión adecuada.

A partir del modelo ER original presentado por Chen (Chen 1976) muchos investigadores han propuesto numerosas extensiones con el objetivo de aumentar su poder de expresión (Elmasri, Weeldreyer et al. 1985; Teorey, Yang et al. 1986; Hohenstein and Gogolla 1988; Czejdo, Elmasri et al. 1990; Markowitz and Shoshani 1992; Chen 2006; Murthy, Delcambre et al. 2006; Combi, Degani et al. 2008). Estas extensiones se han ido incorporando a un supuesto

[1] En la presente tesis se asume una ligera imprecisión en su denominación en español (Entidad-Interrelación) para conservar la sigla ER tan frecuentemente citada.

modelo ER abstracto, que no siempre es el modelo original de Chen; y no es una tarea sencilla unificar todas estas extensiones en un solo modelo, genéricamente denominado modelo Entidad Relación Extendido (EER, acrónimo del inglés Extended Entity Relationship).

Varias de las extensiones del modelo ER original persiguen mayor riqueza semántica en la descripción del universo de discurso. Detrás de las diferencias sintácticas de estas extensiones está el enriquecimiento semántico de las construcciones. Muchos de los cambios sintácticos propuestos están relacionados con la abstracción de generalización y en menor grado con la inclusión de nuevas construcciones. También se puede notar que existen diferencias en cuanto a las notaciones empleadas por diferentes autores en la representación gráfica utilizada en los diagramas (Song, Evans et al. 1995; Hay 1999), para lo cual la notación de Almaro y Navathe (Elmasri and Navathe 2007) se considera un referente adecuado.

Por otra parte, un número creciente de herramientas CASE (acrónimo del inglés Computer Aided Software Engineering) han sido desarrolladas con el objetivo de apoyar a los diseñadores en el proceso de diseño de base de datos. Estas herramientas se caracterizan por una interfaz de usuario que da soporte al modelo conceptual y por un diccionario de datos donde se almacenan los esquemas resultantes. Entre sus beneficios está automatizar el proceso de obtención de un esquema de base de datos, minimizar los costos y el tiempo de implementación, utilizar metodologías y modelos de datos consolidados y contribuir a eliminar errores que el diseñador pueda cometer en las distintas fases del diseño, entre otros aspectos.

Una brecha identificada está relacionada con las capacidades de validación de esquemas de estas herramientas, que en opinión de (Piattini and Díaz 2000) es la más poderosa ayuda que las herramientas CASE deben proveer, ya que tienen un impacto directo en la calidad de los esquemas obtenidos durante el proceso de diseño. Idealmente, el proceso de creación de un esquema ER mediante una herramienta CASE, requiere de un soporte mínimo en términos de la validación de esquemas, que comprende la validación de la sintaxis y la estructura del esquema conceptual.

La validación sintáctica comprende el chequeo de reglas inherentes al modelo ER, como por ejemplo, no permitir ciclos en una jerarquía de generalización y no más de una interrelación débil entre una entidad fuerte y una débil, entre otras. La validación estructural debe comprobar que el conjunto de restricciones de cardinalidad asociado a un esquema conceptual pueda ser satisfecho totalmente, en cuyo caso el esquema se considera consistente. En este sentido se han publicado numerosos trabajos cuyo objetivo es detectar inconsistencias en esquemas conceptuales basados en las restricciones de cardinalidad (Lenzerini and Nobili 1990; Thalheim 1992; Calvanese and Lenzerini 1994; Hartmann 1998; Hartmann 2000; Hartmann 2001; Hartmann 2001; Hartmann 2001; Dullea, Song et al. 2003; Hartmann 2003).

Planteamiento del problema:

El modelo Entidad-Relación (ER) ha sido la base de varias metodologías de diseño para el desarrollo de sistemas de información y de bases de datos relacionales. Un criterio para medir el éxito en el diseño de estos sistemas es el nivel de precisión que ellos reflejan del problema que está siendo modelado. Un modelo puede ser una estructura abstracta y compleja, y los diseñadores están propensos a cometer errores que incorporan inconsistencias. Los errores e inconsistencias cometidas en las etapas iniciales del diseño del ciclo de vida del software son muy costosos de corregir, por lo que es deseable descubrirlos en las etapas iniciales del diseño. Para ayudar a los diseñadores de bases de datos se han desarrollado herramientas comerciales que permiten dibujar los diagramas, las que entre otras facilidades, posibilitan la transformación de esquemas hacia algún formato de bases de datos. Por regla general, estas herramientas antes de realizar las transformaciones realizan un proceso de verificación del esquema conceptual, que generalmente consiste en la determinación de entidades sin nombre, atributos sin tipos asignados, entidades que no estén relacionadas, etc. A pesar de las verificaciones que hacen estas herramientas, un diagrama ER puede presentar inconsistencias que conllevan a que el mismo no sea válido desde el punto de vista estructural. Un diagrama ER es válido estructuralmente cuando todas las restricciones estructurales impuestas en el modelo no implican inconsistencias lógicas de cualquiera de los posibles estados.

En la literatura se reportan varios trabajos que abordan el problema de determinar la consistencia de un conjunto de restricciones de un esquema conceptual, utilizando dos enfoques en su solución: el de programación lineal y el basado en la teoría de grafos. El primero reduce el problema a encontrar una solución a un sistema de inecuaciones lineales. El segundo, permite la detección de restricciones de cardinalidad inconsistentes mediante la búsqueda de ciclos que cumplan con determinadas condiciones. El análisis de estos métodos sienta las bases para su implementación computacional en una herramienta de diseño de bases de datos.

Objetivo general:

Realizar un análisis de los métodos empleados para la validación estructural de esquemas conceptuales Entidad-Relación, con el propósito de recomendar su inclusión en una herramienta de ayuda al diseño de bases de datos relacionales.

Objetivos específicos:

1. Analizar los métodos de validación estructural reportados en la literatura.
2. Definir los algoritmos para cada uno de los métodos de validación estructural.
3. Conformar casos de estudio a los cuales se les serán aplicados los métodos de validación estructural.
4. Seleccionar aquellos métodos de validación de esquemas conceptuales que incluyan una mayor diversidad de construcciones del modelo ER.

Las **preguntas de investigación** planteadas son:

1. ¿De qué manera se pueden expresar las restricciones cardinalidad?
2. ¿Cómo tienen en cuenta los métodos de validación estructural las distintas formas de expresar las restricciones de cardinalidad?
3. ¿Cómo influyen en la eficacia de un método de validación estructural las diversas construcciones del modelo ER?

Después de haber realizado el marco teórico se formula la siguiente **hipótesis de investigación**: en la medida en que los métodos de validación estructural de esquemas conceptuales puedan aplicarse a una amplia variedad de construcciones aumenta la eficacia de los mismos.

1 VALIDACIÓN DE ESQUEMAS CONCEPTUALES ER

El proceso de diseño de una base de datos comienza con la ingeniería de requisitos. Una actividad clave durante este proceso es la creación de un esquema conceptual, el cual describe las propiedades que el sistema de información debe tener y sirve como base para las siguientes fases del diseño. Aunque la modelación conceptual representa sólo una pequeña parte del desarrollo de un sistema, esta tiene un impacto importante en el resultado final; insuficiencias en el diseño conceptual constituyen un factor de peso en el fracaso de un sistema de información (Shanks, Tansley et al. 2003).

Durante la creación del esquema conceptual se debe tener en cuenta chequear la validez estructural y semántica del mismo. Un esquema válido contribuye a mejorar la calidad de los esquemas de bases de datos generados en las siguientes fases del diseño. Por esta razón en este capítulo se hace un estudio sobre los métodos reportados en la literatura que permiten chequear la consistencia de los esquemas conceptuales Entidad-Relación.

1.1 Validación estructural de esquemas conceptuales

En la modelación conceptual, las restricciones de integridad son utilizadas para especificar la manera en que los elementos de una base de datos van a estar asociados entre sí, y por tanto especifican cuáles instancias de la base de datos son válidas (Olivé 2007).

Las restricciones de cardinalidad están entre los tipos de restricciones de integridad más ampliamente utilizadas en la modelación conceptual (Calì 2007). Su función es restringir el número de interrelaciones en que una entidad participa en un tipo de interrelación.

Existen varias maneras de definir la cardinalidad de una interrelación de asociación. Se adopta la definición de cardinalidad para interrelaciones de asociación de grado *n* formulada en (Olivé 2007). Por ejemplo, para interrelaciones binarias:

- $Card(E_1;E_2;R)=(min,\ max)$ indica el número mínimo y máximo de entidades del tipo entidad E_2 que pueden estar interrelacionadas en R, con alguna entidad del tipo entidad E_1.

- $Cmin(p_1{:}E_1,p_2{:}E_2{:}R)$ y $Cmax(p_1{:}E_1,p_2{:}E_2{:}R)$ denotan el valor de la cardinalidad mínima y máxima del tipo de entidad E_1 a través del rol p_1.

Por ejemplo, para el tipo de interrelación Lee (lector: Persona, lectura: Libro) las cardinalidades:

Card(lector: Persona, lectura: Libro; Lee) = (0,3)

Card(lectura: Libro, lector: Persona; Lee) = (0,1)

Significan que una persona puede leer entre 0 y 3 libros en cualquier momento, y que un libro puede ser leído por 0 ó 1 persona en cualquier momento.

El uso de 0 y 1 para la cardinalidad mínima y de 1 y N para la cardinalidad máxima es el enfoque más ampliamente utilizado en (Chen 1976; Markowitz and Shoshani 1992; Date 2003; Ponniah 2003; Silberschatz, Korth et al. 2006; Teorey, Lightstone et al. 2006; Elmasri and Navathe 2007; Ullman and Widom 2007; Garcia-Molina, Ullman et al. 2008). Otros enfoques generalizan las restricciones de cardinalidad de manera que permiten valores mayores que 1 para la cardinalidad mínima y máxima (Lenzerini and Santucci 1983; Lenzerini and Nobili 1990; Thalheim 1992; Calvanese and Lenzerini 1994; Thalheim 1999; Thalheim 2000; Balaban, Maraee et al. 2007). Algunos autores generalizan el concepto de cardinalidad posibilitando que esta puede ser expresada como un conjunto de números $\{min_1,\ldots,max_1,\ldots,min_n,\ldots,max_n\}$ (Hartmann 1998; Hartmann 2000; Hartmann 2001; Hartmann 2001).

El proceso de adquisición de las restricciones de cardinalidad está lejos de ser una tarea trivial (Hartmann, Link et al. 2009), la semántica se recopila a partir de varias fuentes que pudieran generar conflictos. La práctica demuestra que la integración de la información captada puede conducir a obtener restricciones de cardinalidad inconsistentes, por lo que el problema radica en: 1) ¿cómo identificar las inconsistencias?, 2) ¿cómo reportarlas al diseñador?, y 3) ¿cómo corregirlas? (Hartmann 2001).

Sea *S* un esquema ER con un conjunto *C* de restricciones de cardinalidad asociadas. Se dice que un esquema *S* es satisfecho desde el punto de vista estructural si admite al menos una instancia válida de la base de datos. Sin embargo puede suceder que para algunas restricciones de cardinalidad el conjunto de instancias válidas sea vacío o infinito. Se dice entonces que un esquema *S* es totalmente satisfecho si admite un conjunto no vacío y finito de instancias válidas. Esquemas que no sean totalmente satisfechos se consideran incorrectos (Olivé 2007).

Formalmente, un esquema ER *S* con un conjunto *C* de restricciones de cardinalidad es consistente o totalmente satisfecho si existe al menos una base de datos $DB = (r_1, \ldots, r_k)$ acorde a (*S*,*C*) en la cual todas las instancias r_i sean no vacías (Thalheim 1992).

En la literatura se reportan varios trabajos que abordan el problema de determinar la consistencia de un conjunto de restricciones de un esquema conceptual, utilizando dos enfoques en su solución: el de programación lineal y el basado en la teoría de grafos. El primero reduce el problema a encontrar una solución a un sistema de inecuaciones lineales. El segundo, permite la detección de restricciones de cardinalidad inconsistentes mediante la búsqueda de ciclos que cumplan con determinadas condiciones.

1.1.1 Detección de inconsistencias utilizando la programación lineal

Uno de los primeros trabajos relacionados con la satisfacción de las restricciones de cardinalidad en un esquema ER aparece en (Lenzerini and Nobili 1990). El método consiste en transformar las restricciones de cardinalidad en un sistema de inecuaciones lineales, cuyas variables representan los valores de las cardinalidades de los tipos de entidad y tipos de interrelación en una posible instancia. Por ejemplo, en la interrelación que se muestra en la Figura 1.1:

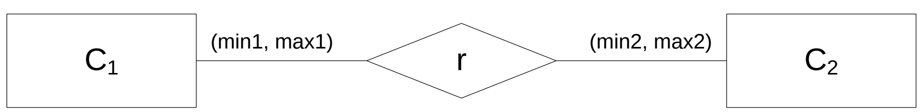


Figura 1.1. Esquema ER con restricciones de cardinalidad expresadas con la notación "look here".

Se obtienen cuatro inecuaciones: $r \geq min_1 \cdot C_1$, $r \leq max_1 \cdot C_1$, $r \geq min_2 \cdot C_2$, $r \leq max_2 \cdot C_2$, donde r, C_1, C_2, son variables que representan la cantidad de instancias de sus respectivos tipos de entidad o tipos de interrelación.

De acuerdo al método, las restricciones de cardinalidad de un diagrama ER pueden ser satisfechas sí y sólo sí el sistema de inecuaciones tiene una solución. Como limitaciones se encuentran que sólo se puede aplicar en esquemas ER que contengan interrelaciones de asociaciones recursivas y binarias, y que no determina cuáles son las restricciones de cardinalidad que hacen que el esquema conceptual sea inconsistente. En (Calvanese and

Lenzerini 1994) se extiende el sistema de inecuaciones basado en el método anterior para aplicarlo a esquemas ER que incluyan interrelaciones de subconjunto. La extensión se basa en el supuesto de que las entidades de los tipos de entidades especializados pueden solaparse. Se proporciona un algoritmo con dos fases, en el cual el problema de la satisfacción de las restricciones de cardinalidad de un diagrama ER, que incluye interrelaciones de subconjunto, se reduce al problema de satisfacción de las restricciones de cardinalidad de un diagrama ER que no contenga interrelaciones de subconjunto. Entonces de manera similar al método anterior se prueba que el sistema de inecuaciones de este nuevo diagrama tenga una solución. Debido a que el sistema de inecuaciones es más complejo que el propuesto en (Lenzerini and Nobili 1990), su complejidad temporal en el peor caso es exponencial. Este método fue simplificado en (Cadoli, Calvanese et al. 2004) restringiendo solamente el solapamiento a las jerarquías de subconjunto, lo cual reduce el número de tipos de entidades y de asociaciones entre ellas, pero la complejidad temporal para el peor caso sigue siendo exponencial.

Debido a la similitud entre las restricciones de cardinalidad en el modelo ER y la multiplicidad en el modelo UML (acrónimo del inglés Unified Modeling Language), los resultados obtenidos en el análisis de las restricciones de cardinalidad han sido utilizados en UML (Balaban and Maraee 2006; Balaban, Maraee et al. 2007), extendiendo el sistema de inecuaciones analizado en (Lenzerini and Nobili 1990) para aplicarlo a diagramas UML que pueden incluir: asociaciones binarias, jerarquías de clases, jerarquías de generalización, asociaciones de orden *n* y clases de asociación. La extensión se basa en un preprocesamiento que reduce el problema de la satisfacción de las restricciones de integridad del diagrama UML al problema de satisfacción de las restricciones de integridad analizado al comienzo de este epígrafe. La ventaja de este método en comparación con el método de Calvanese y Lenzerini (Calvanese and Lenzerini 1994) radica en su simplicidad y eficiencia, lo que lo hace adecuado para su inclusión en una herramienta de diseño basado en UML. Como limitación está que el método no es aplicable cuando existen ciclos en una jerarquía de clases.

1.1.2 Detección de inconsistencias utilizando la teoría de grafos

En (Lenzerini and Nobili 1990), también se propone un método para identificar restricciones de cardinalidad inconsistentes, basado en la representación del esquema ER mediante un multigrafo dirigido y en la búsqueda de ciclos que cumplan con una determinada condición. En el multigrafo dirigido $G = (V, A)$ asociado al esquema conceptual S, el conjunto de vértices V está formado por los tipos de entidades y los tipos de interrelaciones del esquema conceptual. El conjunto de arcos A está determinado por las siguientes reglas: por cada conexión en S entre un tipo entidad E_1 y un tipo de interrelación R a través del rol p_1, se establecen dos arcos e_1 y $e_2 \in A$; e_1 va del nodo E_1 al nodo R y tiene como peso $Cmax(p_1{:}E_1,p_2{:}E_2;R)$; e_2 va del nodo R al nodo E_1 y tiene como peso $\frac{1}{\mathrm{Cmin}(p_1 : E_1, p_2 : E_2; R)}$ si $Cmin(p_1{:}E_1,p_2{:}E_2;R) \neq 0$, ó ∞ si $Cmin(p_1{:}E_1,p_2{:}E_2;R) = 0$. De manera similar se establecen los arcos entre E_2 y R a través del rol p_2. En este trabajo los autores han considerado que $Cmin(p_1{:}E_1,p_2{:}E_2;R)$ y $Cmax(p_1{:}E_1,p_2{:}E_2;R)$ pueden tomar cualquier valor positivo siempre que se cumpla que $Cmin(p_1{:}E_1,p_2{:}E_2;R) \leq Cmax(p_1{:}E_1,p_2{:}E_2;R)$.

De esta manera el diagrama ER de la Figura 1.2 queda representado en un grafo dirigido como se muestra en la Figura 1.3.

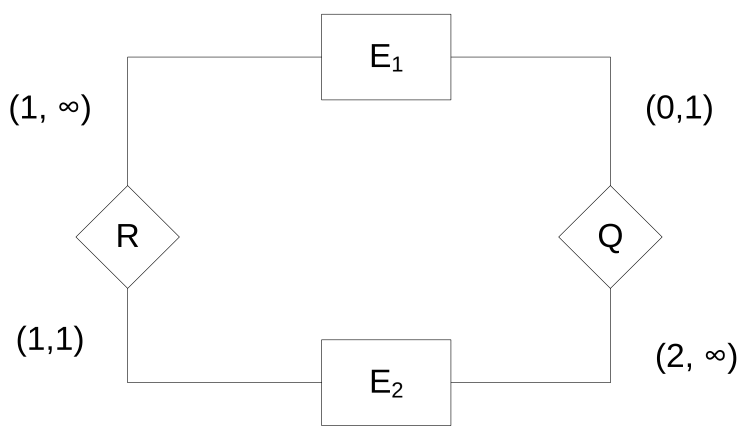


Figura 1.2. Esquema ER con restricciones de cardinalidad.

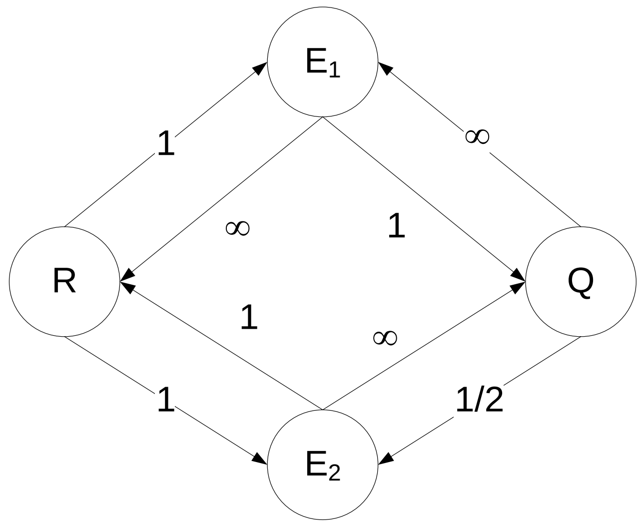


Figura 1.3. Grafo correspondiente al esquema ER de la Figura 1.2.

La determinación de restricciones de cardinalidad inconsistentes se realiza buscando ciclos críticos. Un ciclo crítico es una secuencia de arcos (v_0,v_1) (v_1,v_2) … (v_{k-1},v_k) que debe cumplir con las siguientes condiciones:

1. $v_k = v_0$; $v_1, \ldots , v_k$ son distintos; y
2. el producto del peso de los arcos (v_0,v_1) (v_1,v_2) … (v_{k-1},v_k) es menor que 1.

Cada ciclo crítico encontrado en el grafo se corresponde con un conjunto de restricciones de cardinalidad que no pueden ser satisfechas en el esquema conceptual. El método sólo puede ser aplicado a esquemas que contengan interrelaciones de asociaciones recursivas y binarias.

El método queda descrito de la siguiente manera:

Paso 1. Construir el grafo asociado al diagrama ER.

Paso 2. Hallar los ciclos del grafo.

Paso 3. Multiplicar los pesos de las aristas para cada uno de los ciclos encontrados en el paso anterior.

Una extensión a esquemas que incluyan interrelaciones de orden *n* fue desarrollada en (Thalheim 1992) donde se formaliza una notación de las restricciones de cardinalidad para las interrelaciones de asociación de orden *n*, y se generaliza el concepto de restricciones de cardinalidad, de manera que las restricciones de cardinalidad pueden ser especificadas como un conjunto de números en lugar de un intervalo de valores. La representación del esquema conceptual mediante un grafo es similar a (Lenzerini and Nobili 1990) aunque con una ligera variación para soportar las restricciones de cardinalidad de las interrelaciones de orden *n*. Una vez construido el grafo la detección de las restricciones de cardinalidad inconsistentes sigue el mismo procedimiento analizado en (Lenzerini and Nobili 1990). Para la corrección de esquemas inconsistentes, se propone un algoritmo, el cual toma como entrada un esquema ER formado por los tipos entidades y tipos de interrelaciones, un conjunto de restricciones de cardinalidad asociados al esquema ER y un conjunto formado por todos los ciclos críticos detectados en el grafo. El algoritmo elimina del esquema ER los tipos de entidades y tipos de interrelaciones que forman parte de un ciclo crítico así como todos aquellos tipos de interrelaciones cuyos componentes forman parte de algún ciclo crítico, obteniéndose como resultado un esquema consistente. Nótese que en este método no se hace una corrección de las restricciones de cardinalidad inconsistentes sino que se eliminan del esquema ER los tipos de entidades y tipos de interrelaciones que inciden en la inconsistencia del esquema conceptual.

Otros métodos que abordan el problema de la satisfacción de las restricciones, incluyen además de la detección de inconsistencias, la identificación de las causas y un plan para la corrección del esquema ER. En (Hartmann 2000; Hartmann 2001) se proponen métodos basados en los ciclos críticos para identificar las causas de las inconsistencias y se sugieren cuatro estrategias heurísticas para la corrección de los esquemas conceptuales inconsistentes. (Hartmann 2001). Uno de los métodos está basado en la búsqueda de conjuntos mínimos de restricciones de cardinalidad inconsistentes en cada ciclo crítico. Un conjunto Σ_0 de restricciones de cardinalidad inconsistentes es mínimo si cada subconjunto propio de Σ_0 es consistente, donde $\Sigma_0 \subset \Sigma$ y Σ es el conjunto de restricciones de cardinalidad asociado al esquema conceptual. El procedimiento que se sigue para eliminar las inconsistencias consiste en: para cada subconjunto mínimo Σ_0 se selecciona una restricción de cardinalidad $\sigma \in \Sigma_0$ la cual es eliminada del conjunto Σ. Este método es preciso para la resolución de inconsistencias pero tiene dos potenciales desventajas. Primero, no es posible estimar el tiempo de ejecución ya que no se conoce con qué frecuencia el algoritmo debe ser reiniciado, esto es, cuántos conflictos deben ser tratados. Segundo, la detección de ciclos críticos es un proceso paso a paso, primero se hace la corrección de un conflicto antes de buscar y tratar el próximo. Esto puede ser mejorado posponiendo el proceso de corrección hasta el final del análisis y presentando al diseñador un plan de posibles correcciones, lo cual es logrado a través del método basado en la noción de conjunto de arcos de retroalimentación (*feedback arc set*). Sea G un grafo dirigido, un conjunto F de arcos en G se dice ser de retroalimentación si este intercepta todos los ciclos críticos en G. Un conjunto de arcos de retroalimentación F se obtiene seleccionando un ordenamiento lineal $\underline{v}_1,\ldots,\underline{v}_n$ de los vértices del grafo dirigido G. Entonces, los arcos inversos (vj,vi) donde j > i, forman el conjunto de arcos de retroalimentación. La estrategia de corrección es la siguiente: siempre que una arista l=($\underline{r}$,$\underline{c}$) o su arco inverso l^{-1}=($\underline{c}$,$\underline{r}$) pertenezcan a F, la restricción de cardinalidad Card($\underline{r}$,$\underline{c}$)=M se corrige de la siguiente manera: 1) si l=($\underline{r}$,$\underline{c}$) pertenece F entonces Cmin($\underline{r}$,$\underline{c}$) $\leftarrow$ 0; 2) si l^{-1}=($\underline{c}$,$\underline{r}$) pertenece a F entonces Cmax($\underline{r}$,$\underline{c}$) $\leftarrow \infty$. La complejidad computacional del algoritmo que determina el conjunto de arcos de retroalimentación es NP-hard. Una de las ventajas de este método es la de proporcionar al diseñador de la base de datos un plan de correcciones a aplicar sobre el esquema conceptual. El diseñador debe hacer la selección de los esquemas que deben ser corregidos. Estos resultados fueron extendidos en (Hartmann 2003) de manera que las restricciones de cardinalidad son tratadas como restricciones suaves (*soft constraints*) basado en: un promedio de satisfacción del conjunto de restricciones de cardinalidad, o en un porciento esperado de satisfacción del conjunto de restricciones de cardinalidad. Para ambos enfoques se proponen métodos para la corrección de esquemas, similares a los ya mencionados en (Hartmann 2001). Como ventaja se señala la de proporcionar al diseñador un plan de correcciones a aplicar sobre el esquema conceptual.

Otro método que realiza la detección e identificación de las causas de las inconsistencias en un esquema conceptual ER es el propuesto en (Dullea and Song 1998; Dullea and Song 1998; Dullea and Song 1999; Dullea, Song et al. 2003), el cual se basa en el análisis de las restricciones de cardinalidad y su influencia en las interrelaciones de asociación de grado 1, 2 ó 3 cuando estas forman parte de algún ciclo en el diagrama ER. Los valores utilizados para expresar las restricciones de cardinalidad son: 0 y 1 para la cardinalidad mínima; 1 y N para la cardinalidad máxima.

Para las interrelaciones recursivas se presenta una clasificación de las mismas basada en los roles que una entidad puede desempeñar en la interrelación, lo cual posibilita formular reglas que verifican la consistencia de las restricciones de cardinalidad impuestas en interrelaciones recursivas.

Para el estudio de las interrelaciones binarias se utiliza el concepto de camino (Dullea and Song 1997), que es la conectividad que se establece entre los tipos de entidades y los tipos de interrelaciones en un esquema ER, y muestran que en un camino cíclico cada conjunto de restricciones de cardinalidad debe ser consistente con el resto de las restricciones de cardinalidad y en ese sentido establecen las reglas para determinar la validez estructural de las interrelaciones binarias.

Con respecto a las interrelaciones ternarias, en (Song and Jones 1993; Jones and Song 1996; Jones and Song 2000; Jones and Song 2002) se hace un amplio estudio de la coexistencia de interrelaciones ternarias y binarias en caminos cíclicos, identificándose tres tipos de coexistencias de interrelaciones ternarias y binarias: 1) interrelaciones ternarias no restringidas con interrelaciones binarias implícitas entre los tipos de entidades de la interrelación ternaria; 2) interrelaciones ternarias restringidas por interrelaciones binarias; y 3) interrelaciones ternarias con interrelaciones binarias que no restringen la interrelación. Del estudio se derivan las reglas que determinan la validez estructural de las interrelaciones ternarias. Otros trabajos, como los que se presentan en (Thalheim 1992; McAllister 1998; Santos, Martínez et al. 2007) sólo analizan las interrelaciones ternarias de manera independiente sin tener en cuenta la interacción con otras interrelaciones.

El algoritmo que realiza la validación estructural de un esquema ER basado en el conjunto de reglas ya mencionadas se basa en la representación del esquema conceptual mediante un multígrafo.

Sea G = (V, A) el multígrafo que representa al esquema conceptual S. El conjunto de vértices V está formado por los tipos de entidades y tipos de interrelaciones. El conjunto de arcos A se establece de la siguiente manera:

- Si $R(p_1{:}E, p_2{:}E)$ es un tipo de interrelación recursiva se crean dos arcos e_1 y e_2 entre el tipo de interrelación R y el tipo de entidad E, con pesos $w(e_1) = Card(p_1{:}E; p_2{:}E)$ y $w(e_2) = Card(p_2{:}E; p_1{:}E)$.
- Si $R(p_1{:}E_1, p_2{:}E_2)$ es un tipo de interrelación binaria se crean dos arcos e_1 y e_2 entre el tipo de interrelación R y los tipos de entidades E_1 y E_2, respectivamente, con pesos $w(e_1) = Card(p_1{:}E_1; p_2{:}E_2)$ y $w(e_2) = Card(p_2{:}E_2; p_1{:}E_1)$.
- Si $R(p_1{:}E_1, p_2{:}E_2, p_3{:}E_3)$ es un tipo de interrelación ternaria se crean tres arcos e_1, e_2 y e_3 entre el tipo de interrelación R y los tipos de entidades E_1, E_2 y E_3, respectivamente, con pesos $w(e_1) = Card(p_2, p_3; p_1)$, $w(e_2) = Card(p_1, p_3; p_2)$ y $w(e_3) = Card(p_1, p_2; p_3)$. En este caso la cardinalidad se interpreta como ¿cuántas instancias del último rol están interrelacionadas con un par de instancias de los dos primeros roles? Nótese que los autores han utilizado tres de las doce posibles restricciones de cardinalidad que se pueden especificar para una interrelación ternaria según (McAllister 1998; Thalheim 2000), y que son las más utilizadas en notaciones del modelo ER y por herramientas de diseño de bases de datos.

De acuerdo a lo anterior, el diagrama ER de la Figura 1.4 se representa en un multigrafo como se muestra en la Figura 1.5.

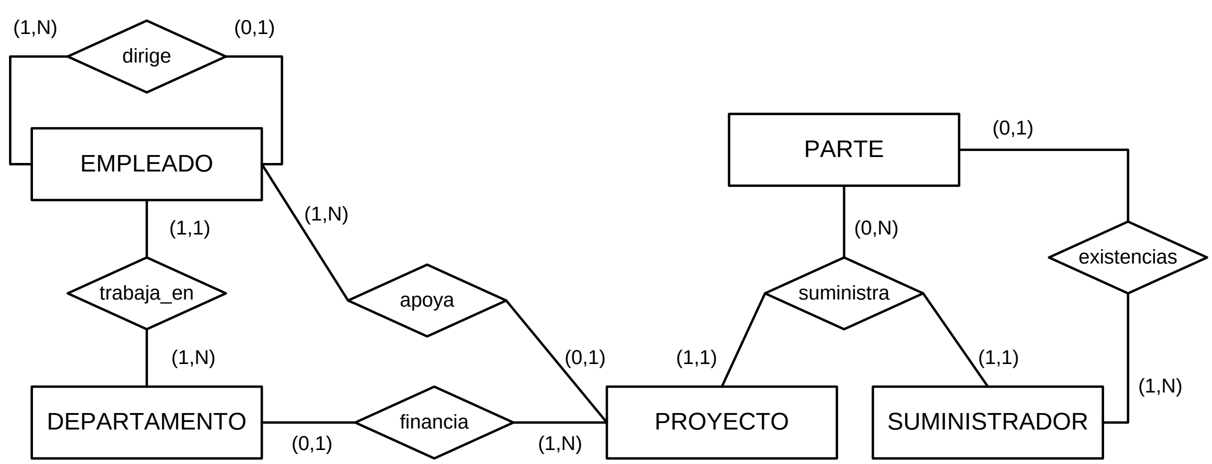


Figura 1.4. Diagrama ER con varios tipos de interrelaciones.

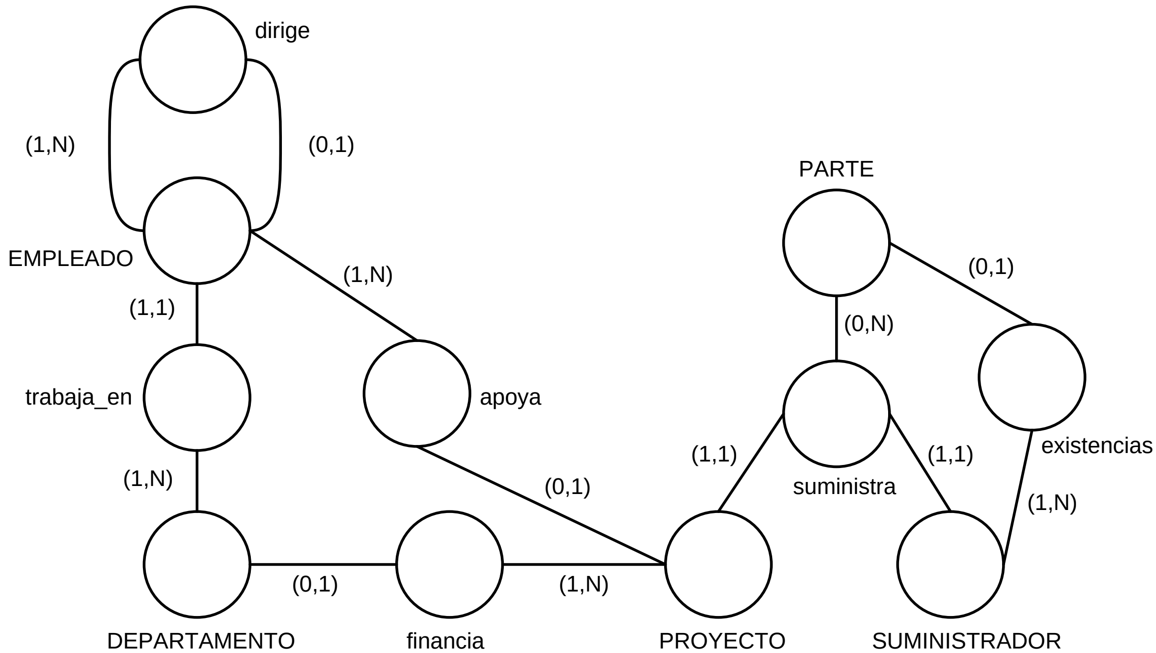


Figura 1.5. Representación en un multígrafo el diagrama ER de la Figura 1.4.

El algoritmo para la evaluación de las reglas es el que se presenta a continuación:

Paso 1. Construir el grafo asociado al diagrama ER.

Paso 2. Para todas las interrelaciones ternarias que tengan interrelaciones binarias impuestas como restricción, determinar la cardinalidad implícita de las interrelaciones binarias derivadas de una interrelación ternaria, las cuales son comparadas con la cardinalidad de las interrelaciones binarias impuestas como restricción para aplicar las reglas correspondientes.

Paso 3. Determinar los caminos cíclicos que forman las interrelaciones binarias en el diagrama. Los caminos cíclicos también incluyen las interrelaciones binarias implícitas derivadas de las interrelaciones ternarias y evaluar las reglas correspondientes.

Paso 4. Para todas las interrelaciones recursivas evaluar las reglas correspondientes.

Haciendo una valoración de los métodos descritos, se puede notar que los basados en la programación lineal sólo determinan si el conjunto de restricciones estructurales impuestas en el esquema conceptual puedan ser cumplidas totalmente, o sea, que no existan instancias vacías en la base de datos. Los métodos que utilizan la teoría de grafos son más precisos que los anteriores, ya que permiten identificar las restricciones de cardinalidad que hacen que el esquema conceptual sea inconsistente.

Entre estos métodos se observan aquellos que utilizan restricciones de cardinalidad con valores mayores que 1 para expresar la cardinalidad mínima y máxima, lo que permite enriquecer la semántica de los esquemas ER; sin embargo este tipo de restricciones de cardinalidad es un subconjunto de las restricciones de cardinalidad menos restrictivas, que son aquellas que utilizan los valores 0 y 1 para la cardinalidad mínima y de 1 y N para la cardinalidad máxima. Al respecto Hartmann (Hartmann, Link et al. 2009) reconoce que “un conjunto de restricciones de cardinalidad restrictivas es consistente siempre y cuando sea también consistente el conjunto de restricciones de cardinalidad menos restrictivas”. En este sentido, el método propuesto en (Dullea, Song et al. 2003) es más general ya que permite detectar los conjuntos de restricciones de cardinalidad inconsistentes que serían detectados con los métodos anteriormente descritos. Por esta razón, se recomienda la implementación de este método en una herramienta CASE.

1.2 Validación semántica de esquemas conceptuales

A diferencia de la validez estructural, la validez semántica es muy subjetiva y compleja de definir normas para la misma (Kolapalli 2008). Un modelo se dice que es semánticamente válido si es exactamente modelado el concepto de intención del modelador y la comunidad de usuarios, y está libre de cualquier defecto que haría inaplicable, incompleta, o representa la abstracción errónea del concepto de diseño. Es el grado de conformidad del modelo con los requisitos establecidos e implícitos. Según Kolapalli (Kolapalli 2008) los requisitos se clasifican en:

- Requisitos funcionales.
- Requisitos no funcionales.

Los requisitos funcionales definen el objetivo principal del sistema o qué es lo que se supone que debe hacer el sistema. Los requisitos no funcionales, también llamados comúnmente como atributos de calidad o metas, definen el comportamiento del sistema en el logro de su propósito. Un modelo debe satisfacer todos los factores funcionales y cumplir con los factores de calidad en al menos un valor umbral para ser considerado como semánticamente válido.

La validación semántica es muy subjetiva, a continuación se muestran métricas establecidas en (Kolapalli 2008) para evaluar de alguna manera los dos factores mencionados anteriormente:

Factores funcionales

- Integridad
- Corrección
- Coherencia

La integridad de un modelo de datos es el concepto de modelación o de representación de toda la información necesaria para apoyar las necesidades de información de los usuarios, las aplicaciones utilizadas por los usuarios, y todos los datos en la base de datos que necesitan de otros datos o mecanismos que permitan a su vez intercambiar conceptos dentro de la información. La integridad asegura que el modelo de datos esté direccionado a los requisitos previstos, mientras que la corrección asegura que los requisitos que se modelan estén libres de errores y que proporcionen los resultados correctos a la hora de implementar.

Además de la corrección semántica en atributos y del modelado de dependencias funcionales, la corrección también valida que ambos tengan la misma semántica especificada en los requerimientos. Las restricciones de participación especifican si todas las instancias de la entidad participan en la relación o no. Si la entidad participa totalmente en la relación, entonces la restricción de participación es obligatoria. Si la entidad tiene una participación parcial en la relación, la restricción de participación es opcional.

La coherencia es una característica de diseño e implementación. La inconsistencia en la base de datos se refiere a un estado en el que una parte de la base de datos no está de acuerdo con otra parte de la base de datos. Esto puede ser debido a las restricciones en conflicto o de nombres incompatibles, dando lugar a respuestas incorrectas en las consultas. La inconsistencia semántica se refiere a tener el mismo nombre para datos que representan campos de datos con significados diferentes. Inconsistencia de nomenclatura se refiere a tener diferentes nombres para elementos que representan campos de datos iguales.

Factores de calidad semántica

- Nivel de abstracción.
- Flexibilidad.
- Integración.

Un nivel adecuado de abstracción (Kolapalli 2008) en el diseño de un modelo de datos es deseado, como menor nivel de abstracción o abstracción excesiva pueden causar efectos adversos. Una estructura más abstracta hace el modelo más preciso y flexible.

Los niveles de abstracción se pueden ver de dos maneras diferentes: La profundidad de la abstracción y la amplitud de la abstracción. La profundidad de la abstracción es el nivel de granularidad de las unidades de datos que se almacenan en la base de datos. Muy poca granularidad limitará la cantidad de información disponible para la toma de decisiones, y probablemente, la falta de capacidad de tomar decisiones de calidad. Demasiada granularidad afectará el costo de almacenar la información innecesaria y también a las consultas de rendimiento de las aplicaciones de acceso a los datos (Kolapalli 2008).

La amplitud no es más que el tamaño del nivel de abstracción. El problema se produce cuando se excluye información que pueda ser necesaria para otros usuarios o cuando se incluye información adicional demasiado grande y que rara vez es utilizada.

La facilidad con que el modelo de datos puede hacer frente a los cambios se conoce como flexibilidad. Un modelo de datos de alta calidad no debería requerir un cambio estructural importante para dar cabida a un pequeño cambio en las reglas del negocio (Kolapalli 2008).

La integración es la actividad de la combinación de componentes en conjunto que provienen de fuentes diferentes para que puedan trabajar al unísono para proporcionar información adicional que ninguno de los componentes puede suministrar por sí solo (Kolapalli 2008). La integración proporciona la flexibilidad necesaria para diseñar y construir componentes de un sistema separado permitiendo concentración individual de las características específicas de cada componente, mientras que el resultado final es trabajar juntos como diferentes instrumentos musicales en una orquesta sinfónica.

1.3 Conclusiones parciales

De lo abordado en este capítulo se puede concluir que:

1. La validación de esquemas conceptuales permite detectar errores e inconsistencias, que de no ser detectadas a tiempo, estas se pueden propagar a otras fases del diseño de la base de datos.
2. Existen varios métodos para realizar una validación estructural de esquemas conceptuales, en este trabajo se han seleccionado dos de ellos para su estudio.
3. La validación semántica es dependiente del problema y las reglas que se pueden establecer dependen en gran medida del problema y del criterio del diseñador. En este trabajo se ha mencionado este aspecto como parte de un proceso integral de validación de esquemas conceptuales pero no se analizarán métodos de validación semántica.

2 ANÁLISIS DE MÉTODOS DE VALIDACIÓN ESTRUCTURAL DE ESQUEMAS CONCEPTUALES ER

Detectar inconsistencias en un diagrama ER durante la ingeniería de requisitos evita insuficiencias en el desarrollo de la base de datos. Con el objetivo de detectar dichas inconsistencias a través de este capítulo se realiza un análisis de dos métodos: el de Lenzerini y Nobili (Lenzerini and Nobili 1990) basado en la multiplicación del peso de las aristas y el de Dullea y Song (Dullea, Song et al. 2003) el cual consiste en un grupo de reglas que determinan la validez del esquema.

2.1 Método de Dullea y Song

En este método se realiza un análisis de las interrelaciones recursivas, binarias y ternarias de un esquema ER. Se demuestra que tanto las relaciones binarias como ternarias cuando son independientes son estructuras válidas, ambas se evalúan de manera integral con respecto a las demás interrelaciones del diagrama dado que se relacionan con otros tipos de entidades. Se refleja el papel importante jugado por las interrelaciones binarias restrictivas en la evaluación de las interrelaciones ternarias.

2.1.1 Interrelaciones recursivas

Una interrelación recursiva es definida como una asociación entre las instancias que asumen diferentes roles sobre una misma entidad (Dullea and Song 1999; Dullea, Song et al. 2003). Una interrelación recursiva es válida estructuralmente cuando la cardinalidad y las restricciones de participación soportan instancias válidas. Según la literatura, las interrelaciones recursivas se clasifican en simétricas o asimétricas. En las interrelaciones simétricas o reflexivas la interrelación toma un sólo rol y el significado semántico de la relación es el mismo desde donde se mire. Las relaciones asimétricas o no-reflexivas son aquellas donde la asociación es entre instancias con dos grupos de roles diferentes sobre la misma entidad y tiene diferente significado dependiendo desde el punto de vista de donde se mire, estas relaciones no pueden ser obligatorias completamente. Las interrelaciones recursivas asimétricas se dividen según (Dullea and Song 1998) en tres tipos: jerárquicas, circular y espejo. Una interrelación recursiva es jerárquica cuando un grupo de instancias dentro de la misma entidad son ordenadas en grados, orden, o clases, una encima de otra. Una interrelación recursiva es circular cuando tiene al menos una instancia que no cumple con el rango de clasificación, este tipo de interrelación es unidireccional ya que no puede ser vista desde direcciones diferentes y tener el mismo significado (Dullea and Song 1999). Una relación es espejo cuando semánticamente se permite que una instancia de una entidad se asocie consigo misma a través de la relación.

A continuación se realiza un análisis de los conjuntos de restricciones de cardinalidad de las interrelaciones recursivas para demostrar si son válidas o no.

Interrelaciones Recursivas 1:1

Obligatorio-Obligatorio: cada instancia del grupo de roles debe participar completamente en la relación. Cuando la relación es 1:1 cada instancia debe ser pareja de una y sólo una instancia. Ejemplo: cada persona de un grupo debe estar casada con sólo una persona del grupo.

Opcional-Opcional: Cada instancia del grupo de roles puede participar opcionalmente en la relación.

Obligatorio-Opcional u Opcional-Obligatorio: cada instancia del grupo de roles debe participar en la relación mientras la instancia del otro grupo puede participar opcionalmente.

Teorema 1: *Una relación recursiva 1:1 con una cardinalidad mínima de Obligatorio-Opcional es una estructura inválida.*

Prueba: Suponga que un grupo de roles G_1 contiene L_j instancias donde j=1, 2,3,…, n y representa todas las instancias del tipo de entidad E.

$|G_1|=|E|=n$ (1)

Debido a que la participación es obligatoria, cada instancia en G_1 se asigna totalmente al grupo de roles 2 (G_2) y $|G_2|$ es por lo menos igual a n. Por lo tanto

$|G_2| \geq n$ (2)

Si las instancias en G_2 son sólo parcialmente asignadas a G_1 entonces el grupo de roles G_2 puede contener al menos una instancia que no es asignada a instancias de G_1. Por lo tanto en su forma estricta queda:

$| G_2| > n$ (3)

G_2 representa todas las instancias en la entidad E por tanto de la ecuación (3) se puede derivar

$|E| > n$ (4)

Como se puede apreciar la ecuación (1) y (4) son inconsistentes porque |E| no puede ser mayor e igual a n al mismo tiempo. Por tanto como queda demostrado la estructura es inválida.

Regla 1: *Solo las estructuras de interrelaciones recursivas 1:1 con Obligatorio-Obligatorio u Opcional-Opcional de mínima restricción de cardinalidad son válidas.*

Corolario 1. *Todas las relaciones recursivas 1:1 con Obligatorio-Opcional u Opcional-Obligatorio de mínima restricción de cardinalidad son estructuras inválidas.*

Interrelaciones Recursivas 1:M

Son siempre asimétricas porque requieren de dos grupos de roles.

Obligatorio-Obligatorio: Es inválido como se demuestra a continuación.

Teorema 2: *Una interrelación recursiva 1:M donde la restricción de cardinalidad mínima es obligatoria es una estructura inválida*.

Prueba: Dado que los roles 1 y 2 están contenidos en el tipo de entidad E, participan completamente en la interrelación R. Como la restricción 1:M está situada sobre R, la interrelación R debe soportar el caso que G_1 contiene menos instancias que G_2, por tanto, en la forma estricta queda:

$|G_1| < |G_2|$ (1)

El hecho de que la restricción sea obligatoria en R implica que las instancias en G_1 representen todas las instancias en E, por tanto

$|G_1|=|E|$ (2)

Entonces el hecho de que la restricción sea obligatoria en R implica que las instancias en G_2 representen todas las instancias en E, por tanto

$|G_2| = |E|$ (3)

Si $|G_1|=|E|$ y $|G_2| = |E|$

Entonces $|G_1|=|G_2|$ (4)

Se tiene que la ecuación (1) y (4) son inconsistentes, por tanto queda demostrado que la estructura es inválida.

Obligatorio-Opcional: es una estructura inválida como se demuestra a continuación.

Teorema 3: *Una interrelación recursiva 1:M donde la restricción de cardinalidad mínima es obligatoria en el lado del 1 y opcional en el lado mucho es una estructura inválida.*

Prueba: Se tienen dos grupos de roles 1 y 2 que están contenidos en el tipo de entidad E y participan en la interrelación R 1:M. Si el G_1 en el lado 1 de la relación 1:M está conectada con la participación opcional del grupo de roles 2 en el lado muchos, entonces la interrelación R debe soportar el caso que G_1 contiene menos instancias que G_2. Por tanto

$|G_1| < |G_2|$ (1)

Como la restricción de cardinalidad es obligatoria en el grupo de roles 1 el número de instancias en G1 debe ser igual al número de instancias en E. Por tanto

$|G_1|=|E|$ (2)

Si se sustituye la ecuación (1) con la (2) se tiene que $|E| < |G_2|$, y se obtiene una inconsistencia porque un grupo de roles no puede tener mayor número de instancias que la entidad donde está contenido el mismo.

Opcional-Opcional: son estructuras válidas.

Regla 2: *Las interrelaciones recursivas con restricciones de cardinalidad 1:M o M:1 y cardinalidad mínima opcional-opcional son estructuras válidas.*

Regla 3: *Las interrelaciones recursivas 1:M del tipo jerárquica circular con cardinalidad mínima opcional-obligatorio son estructuras válidas.*

Corolario 2: *Todas las restricciones recursivas 1:M o M:1 donde las restricciones de cardinalidad mínimas son obligatorio-obligatorio son estructuras inválidas.*

Corolario 3: *Todas las interrelaciones recursivas 1:M o M:1 que presenten obligatoria participación en el lado "uno" y opcional participación en el lado "muchos" son estructuras inválidas.*

Interrelaciones Recursivas M:M

Existen tres casos de estas interrelaciones que deben ser investigadas, a continuación se muestran.

Obligatorio-Obligatorio: Las interrelaciones recursivas "obligatorio-obligatorio" con restricción de cardinalidad M: M son válidas para las relaciones simétricas, mientras que para las asimétricas son inválidas como se demostró anteriormente.

Obligatorio-Opcional: Este tipo de interrelación es válida.

Opcional-Opcional: Para comprender mejor este caso, suponga el ejemplo de un tipo de Persona compuesta por grupos de hermanos y hermanas con "Hermano-de" como interrelación, y si se le adicionan individuos al tipo de entidad que no tienen ni hermanos ni hermanas, entonces se está en presencia de una relación simétrica M:M opcional a ambos lados y completamente válida.

Regla 4: *Todas las interrelaciones recursivas con máxima cardinalidad M:M son estructuralmente válidas sin tener en cuenta la mínima restricción de cardinalidad.*

Regla 5: *Todas las interrelaciones recursivas con mínima cardinalidad opcional-opcional son estructuralmente válidas.*

La Tabla 1 muestra un resumen de las interrelaciones recursivas válidas con algunos ejemplos correspondientes de acuerdo a las restricciones de cardinalidad.

Tipo de Relación	Dirección de la relación	Restricciones de Participación	Restricciones de Cardinalidad	Ejemplo de	
				Relaciones	Roles

Simétrica o Reflexiva	Jerárquico Bidireccional	Opcional-Opcional Obligatorio-Obligatorio	1:1 M:N	Esposa de	Persona
Asimétrica	Circular Unidireccional	Opcional-Opcional Opcional-Opcional Opcional-Obligatorio Obligatorio-Obligatorio	1:M 1:1 M:N	Supervisor Supervisado por.	Jefe-Empleado
	Jerárquica circular Unidireccional	Opcional-Opcional Obligatorio-Obligatorio	1:1		
	Unidireccional	Opcional-Opcional Opcional-Obligatorio Opcional-Opcional Opcional-Obligatorio Obligatorio-Obligatorio	1:M M:N	Supervisor Supervisado por.	Jefe-Empleado
	Espejo Unidireccional	Opcional-Opcional	1:1		

Tabla1: Tipos de interrelaciones recursivas válidas.

2.1.2 Interrelaciones binarias

Una interrelación binaria es una asociación entre el grupo de roles de un tipo de entidad y el grupo de roles de otro tipo de entidad. De acuerdo al número de instancias entre las entidades en una interrelación las entidades pueden ser representadas como: $|E_1| = |E_2|$,

$|E_1| > |E_2|$ o $|E_1| < |E_2|$ y para determinar estos posibles valores se debe tener en cuenta la mínima y máxima cardinalidad de la relación. Por ejemplo en una interrelación binaria 1:1 obligatoria-obligatoria el número de instancias en cada entidad participante en la restricción debe ser siempre igual por definición de la restricción de cardinalidad, mientras que si la interrelación continúa siendo 1:1 pero opcional-opcional el número de las instancias entre las entidades pueden ser igual, menor o mayor dependiendo del número de instancias que participan o no en la interrelación. Cuando se está en presencia (Dullea, Song et al. 2003) de una ruta cíclica los valores relativos de las instancias de cada entidad se ajustan a sí mismo formando toda la ruta, similarmente ocurre con las entidades que son restringidas, en una interrelación M:1 obligatoria-obligatoria la entidad en el lado "*uno*" se acomoda a la interrelación por tener menos instancias que la entidad en el lado "*muchos*".

Según (Dullea and Song 1998) existen 16 posibles combinaciones para una ruta simple formada por dos entidades y una interrelación entre ellas. En la siguiente tabla se muestran estas posibles combinaciones y las restricciones asociadas. Una ruta será estructuralmente inválida sólo cuando no puede soportar todas las restricciones estructurales impuestas en la ruta simultáneamente (Dullea, Song et al. 2003). Para el análisis de las relaciones binarias se abordaran dos tipos de rutas: acíclicas y cíclicas.

Casos	Máxima cardinalidad	Mínima cardinalidad	Restricciones sobre el número de instancias entre entidades
B_1	1:1	Obligatorio-Obligatorio	$\|E_1\| = \|E_2\|$
B_2	1:1	Obligatorio-Opcional	$\|E_1\| < \|E_2\|$
B_3	1:1	Opcional-Obligatorio	$\|E_1\| > \|E_2\|$
B_4	1:1	Opcional- Opcional	Autoajustable
B_5	M:1	Obligatorio-Obligatorio	$\|E_1\| > \|E_2\|$
B_6	M:1	Obligatorio-Opcional	No-restrictivas
B_7	M:1	Opcional-Obligatorio	$\|E_1\| > \|E_2\|$
B_8	M:1	Opcional- Opcional	Autoajustable
B_9	1:M	Obligatorio-Obligatorio	$\|E_1\| < \|E_2\|$
B_{10}	1:M	Obligatorio-Opcional	$\|E_1\| < \|E_2\|$

B_{11}	1:M	Opcional-Obligatorio	Autoajustable
B_{12}	1:M	Opcional- Opcional	Autoajustable
B_{13}	M:N	Obligatorio-Obligatorio	Autoajustable
B_{14}	M:N	Obligatorio-Opcional	Autoajustable
B_{15}	M:N	Opcional-Obligatorio	Autoajustable
B_{16}	M:N	Opcional- Opcional	Autoajustable

Tabla 2. Restricciones estructurales de una interrelación binaria simple.

Rutas acíclicas

Las rutas acíclicas son aquellas que no regresan con una entidad previa, son rutas abiertas donde cada entidad sólo tiene un efecto estructural directo con una entidad adyacente (Dullea, Song et al. 2003). En la figura 2.1 se muestra un ejemplo de una ruta acíclica binaria simple entre los tipos de entidades E_1 y E_2 a través de la interrelación R_1. Como se puede observar a través de la Tabla 2 este tipo de ruta siempre es válida. Si se le agrega otra entidad E_3 asociada a E_2 a través de la interrelación R_2, entonces E_2 estaría asociada a dos entidades y aun así son independientes cada una. Cualquier interrelación puede tomar alguna de las 16 combinaciones de la Tabla 2 sin afectar la validez estructural de la otra interrelación. Toda interrelación binaria simple es siempre estructuralmente válida y si se encuentra relacionada con otras interrelaciones binarias formando una ruta acíclica son estructuralmente independientes y válidas a la vez.

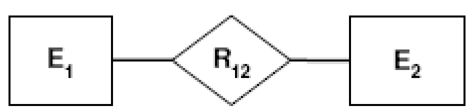


Figura 2.1. Ejemplo de una ruta acíclica.

Regla 6: *Una ruta acíclica contenida completa de interrelaciones binarias es siempre estructuralmente válida*.

Interrelaciones autoajustables y rutas cíclicas

Una ruta cíclica es una ruta cerrada que comienza y termina sobre la misma entidad. Todas las entidades que conforman la ruta están asociadas indirectamente aunque no sean adyacentes lo que implica que existe cierta dependencia entre cada una de ellas.

La Tabla 3 muestra cuatro posibles condiciones que pueden existir para una interrelación binaria derivada de la Tabla 2 y desarrollada a través de un conjunto de heurísticas para determinar la validez estructural de una ruta cíclica que se ajustan a sí misma (Dullea, Song et al. 2003).

Casos	**Restricciones sobre el número de instancias entre entidades**
B_4, B_6, B_8, B_{11}, B_{12}, B_{13}, B_{14}, B_{15}, B_{16}	Autoajustables
B_1	$\lvert E_1 \rvert = \lvert E_2 \rvert$
B_2, B_9, B_{10}	$\lvert E_1 \rvert < \lvert E_2 \rvert$
B_3, B_5, B_7	$\lvert E_1 \rvert > \lvert E_2 \rvert$

Tabla 3. Cuatro posibles condiciones para una relación binaria.

Teorema 4: *Cualquier ruta cíclica de interrelaciones binarias que contiene relaciones del tipo "autoajustables" es estructuralmente válida.*

Prueba: Dada una ruta cíclica que contiene entidades E_1, E_2, E_3,.., $E_{(N-1)}$, E_N con interrelaciones R_{12}, R_{23},..$R_{(N-1)}$, R_N es ajustada a sí misma. La ruta cíclica entre E_1, E_2,..$E_{(N-1)}$, E_N es siempre válida como está representada en la Regla 5 y estará representada por ecuaciones o inecuaciones entre el número de instancias desde E_1 hasta E_N. La interrelación binaria R_{N-1} se ajusta a sí misma así como la interrelación binaria simple entre E_N y E_1 se acomoda a las restricciones que dan lugar a E_N y E_1 mediante la ruta acíclica E_1, E_2, E_3,…, $E_{(N-1)}$, E_N.

Regla 7: *Una ruta cíclica contenida sólo por interrelaciones binarias y una o más interrelaciones "opcional-opcional" es siempre estructuralmente válida.*

Regla 8: *Una ruta cíclica contenida sólo por interrelaciones binarias y una o más interrelaciones M:1 con opcional participación en el lado "uno" es siempre válida.*

Regla 9: *Una ruta cíclica contenida sólo por interrelaciones binarias y una o más interrelaciones M:M es siempre estructuralmente válida.*

Interrelaciones de oposición

En ausencia de una relación "Autoajustable" en una ruta cíclica, cada interrelación en la ruta debe tomar una y sólo una ecuación de condición, $|E_1| = |E_2|$, $|E_1| > |E_2|$ o $|E_1| < |E_2|$ (Dullea, Song et al. 2003). La ecuación de igualdad $|E_1| = |E_2|$ es neutra, por lo tanto no afecta las demás relaciones. La Figura 2.2 muestra un ejemplo de una interrelación de oposición.

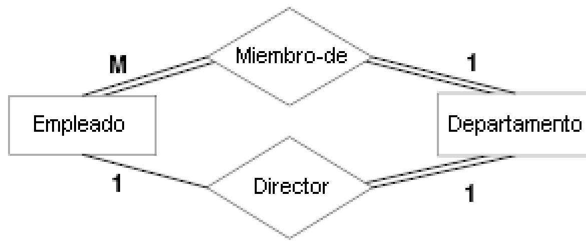


Figura 2.2. Ejemplo de interrelaciones de oposición.

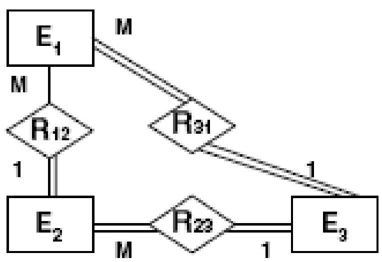


Figura 2.3. Estructura válida.

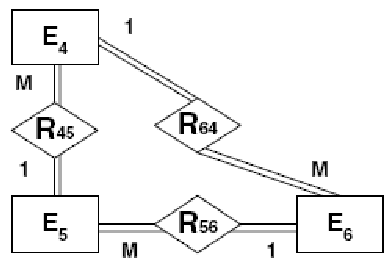


Figura 2.4. Estructura inválida.

Teorema 5: *En la ausencia de una relación "*Autoajustable *"o del tipo* $|E_1| = |E_2|$ *en una ruta cíclica debe existir al menos un conjunto de relaciones de oposición para que la ruta sea válida.*

Prueba: Considérese que la Figura 2.3 representa una estructura válida y contiene un conjunto de instancias válidas para cada entidad. De la Tabla 2 se observa que R_{12} corresponde al caso B_7 y R_{23} al caso B_5

R_{12} de B_7 $\quad |E_1| > |E_2| \quad$ (1)

R_{23} de B_5 $\quad |E_2| > |E_3| \quad$ (2)

Lo que quiere decir que $|E_1| > |E3|$ (3)

De la Tabla 2 R_{31} corresponde al caso B_9

R_{31} de B_9 $\quad |E3| < |E_1|$ (4)

La inecuación (3) es adversario de la (4), lo que quiere decir que el diagrama es una estructura válida.

Prueba: Considérese que la Figura 2.4 es una estructura inválida porque no puede soportar al menos un conjunto de instancias válidas para cada entidad. Desde la Tabla 2 se tiene que:

R_{45} de B_5 $\quad |E_4| > |E_5|$ (5)

R_{56} de B_5 $\quad |E_5| > |E_6|$ (6)

Lo que quiere decir que $\quad |E_4| > |E_6|$ (7)

De la Tabla2 se puede ver que R_{64} corresponde al caso B_5

R_{64} de B_5 $\quad |E_6| > |E_4|$ (8)

Existe una contradicción entre las ecuaciones (7) y (8) por lo que se puede decir que se está en presencia de una ruta cíclica que no contiene relaciones de oposición, esta estructura no puede soportar ningún conjunto de instancias siendo así una estructura inválida.

Regla 10: *Una ruta cíclica que contenga al menos un conjunto de interrelaciones de oposición es siempre válida.*

Corolario 4: *Una ruta cíclica que no contenga interrelaciones de oposición ni interrelaciones autoajustables es estructuralmente inválida y llamada Relación Circular.*

Efecto Neutral

Una relación 1:1 obligatorio-obligatorio (Caso B_1) en una ruta cíclica con más de dos interrelaciones no tiene efecto para las otras interrelaciones y validación de la ruta.

Corolario 5: *La presencia de una interrelación 1:1 obligatorio-obligatorio no tiene efecto sobre la validación estructural (o invalidación) de una ruta cíclica que contiene otros tipos de relaciones. (Este corolario se aplica a todas las reglas posteriores.)*

Regla 11: *Una ruta cíclica contenida completa por interrelaciones binarias 1:1 obligatoria-obligatoria es estructuralmente válida.*

2.1.3 Interrelaciones ternarias

Las interrelaciones ternarias son asociaciones entre tres tipos de entidades, lo que significa que una instancia de una entidad está asociada a dos instancias de las otras dos entidades respectivamente.

Interrelaciones ternarias en una ruta acíclica

Cuando una ruta acíclica está formada por interrelaciones ternarias es estructuralmente válida de manera similar a las interrelaciones binarias porque a la hora de hacer la validación se realiza mediante las interrelaciones binarias que la forman. La interrelación binaria entre dos entidades de una interrelación ternaria puede ser explícita o implícita.

Regla 6 *(actualizada)**:** Si una ruta contenida por interrelaciones binarias y ternarias es una ruta acíclica entonces la ruta siempre es estructuralmente válida.*

Interrelaciones ternarias en una ruta cíclica

En (Dullea and Song 1998) se identifican tres tipos de interrelaciones de coexistencia de interrelaciones ternarias y binarias: interrelaciones ternarias sin restricciones con derivadas relaciones binarias entre entidades, relaciones ternarias más restringidas por interrelaciones binarias explícitas, e interrelaciones binarias sin interrelación entre dos entidades de una relación ternaria.

Interrelaciones binarias no explícitas entre entidades (restringidas o sin relación).

Si una interrelación ternaria no está restringida por una interrelación binaria explícita, entonces tiene una interrelación binaria implícita que describe la asociación entre instancias entre las dos entidades. En (Song and Jones 1993; Jones and Song 1996) se establece una regla de cardinalidad implícita (IBC):

"En cualquier interrelación, sin tener en cuenta la cardinalidad ternaria, la cardinalidad implícita entre dos entidades cualesquiera debe ser considerada M:N, siempre que ellas no sean relaciones explícitas sobre el número de instancias que se producen".

Regla 12: *Una ruta cíclica que contiene una interrelación ternaria que no es restringida explícitamente por interrelaciones binarias sobre las entidades ternarias siempre es* *estructuralmente válida, sin tener en cuenta la máxima y mínima restricción de cardinalidad en la relación ternaria*.

Efecto de las interrelaciones binarias explícitas "sin relación" en una interrelación ternaria.

Cuando la interrelación binaria impone un concepto diferente desde el concepto de ser presentada por la ternaria y no limita las instancias participantes en la interrelación ternaria es considerada "sin relación" (Dullea, Song et al. 2003). La cardinalidad de la interrelación binaria es completamente independiente a la ternaria.

Efecto de interrelaciones binarias explícitas que restringen las interrelaciones ternarias.

Las interrelaciones binarias explícitas se usan para modelar situaciones donde las interrelaciones ternarias no son suficientes y se necesitan para definir aún más la asociación entre dos entidades, forman parte de la interrelación ternaria, en la siguiente figura se muestra un ejemplo de este tipo de relación. En (Jones and Song 2000) se establece un permiso binario explícito (EBP) que plantea que: "*Para cualquier interrelación ternaria dada, una interrelación binaria no puede ser impuesta donde la cardinalidad binaria es menor que la cardinalidad especificada por la ternaria, para cualquier entidad especificada*". Este tipo de interrelación restringe el grupo de instancias de la interrelación ternaria. La cardinalidad implícita entre dos entidades es *"muchos-muchos"* pero puede ser modelado usando la combinación de una interrelación ternaria con una relación binaria restrictiva.

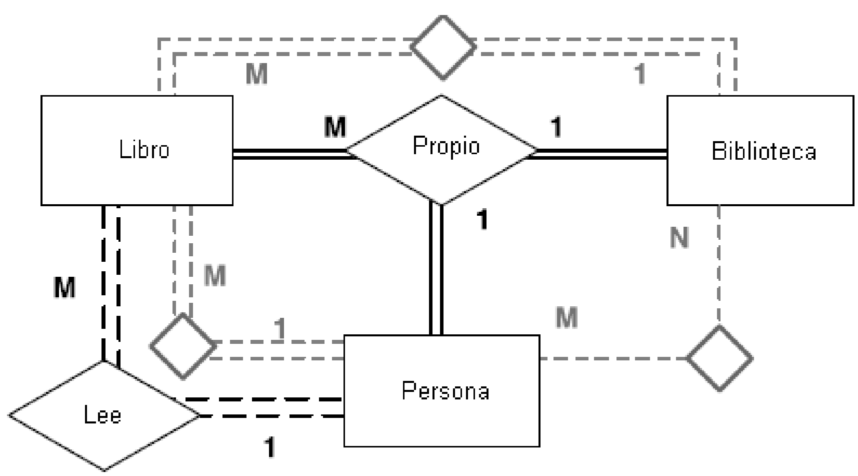


Figura 2.5. Interrelación ternaria restringida por una interrelación binaria explícita.

La siguiente tabla muestra un resumen de las imposiciones binarias permisibles y no permisibles sobre los diferentes grupos de cardinalidad de una relación ternaria.

Interrelación ternaria 1:1:1	Cualquier cardinalidad de una interrelación binaria puede ser impuesta sobre este tipo de interrelación ternaria.

Interrelación ternaria 1: 1: M

X	Y	Z	Permisible
M	N		Si
M	1		Si
1	1		Si
1	M		Si
M		N	Si
M		1	No
1		1	No
1		M	Si
	M	N	Si
	M	1	No
	1	1	No
	1	M	Si

Interrelación ternaria 1: M: N

X	Y	Z	Permisible
M	N		Si
M	1		No
1	1		No
1	M		Si
M		N	Si
M		1	No
1		1	No
1		M	Si
	M	N	Si

	M	1	No
	1	1	No
	1	M	No
Interrelación ternaria M:N:P	Solo la cardinalidad binaria M: N puede ser impuesta a este tipo de interrelación y es redundante ya que la interrelación ternaria implícitamente establece esta restricción de cardinalidad.		

Tabla 4. Restricciones binarias impuestas permisibles o no.

Es importante analizar los cambios que ocurren a las interrelaciones binarias integradas debido a las imposiciones. La tabla siguiente muestra los resultados obtenidos por Jones (Jones and Song 1996) respecto al efecto causado por una interrelación binaria restrictiva impuesta en una relación ternaria.

Interrelación ternaria X:Y:Z	Restricción binaria impuesta	Efecto sobre las interrelaciones integradas
1:1:1	X: Y es M: 1	X: Y es M: 1
		X: Z es M: 1
		Y: Z es M: N
	X: Y es 1: M	X: Y es 1: M
		X: Z es M: N
		Y: Z es M: 1
	X: Y es 1: 1	X: Y es 1: 1
		X: Z es M: 1
		Y: Z es M: 1
M:1:1	X: Y es M: 1	X: Y es M: 1
		X: Z es M: 1
		Y: Z es M: N
	X: Z es M: 1	X: Z es M: 1
		X: Y es M: 1
		Y: Z es M: N
	Y: Z es M: 1	X: Y es M: N
		X: Z es M: 1
		Y: Z es M: 1
	Y: Z es 1: M	X: Y es M: N
		X: Z es M: N
		Y: Z es 1: M

	Y: Z es 1: 1	X: Y es M: N X: Z es M: N Y: Z es 1: 1
M: N: 1	X: Z es M: 1	X: Y es M: N X: Z es M: 1 Y: Z es M: N
	Y: Z es M: 1	X: Y es M: N X: Z es M: N Y: Z es M: 1

Tabla 5. Efectos de una imposición binaria sobre una interrelación ternaria.

Regla 13: *Si la máxima restricción de cardinalidad de la interrelación binaria impuesta sobre una interrelación ternaria es mayor o igual a la máxima restricción de cardinalidad entre las dos entidades implicadas entonces las restricciones de la interrelación ternaria son válidas.*

Corolario 6: *Si la restricción de cardinalidad máxima para una interrelación binaria restrictiva impuesta sobre una interrelación binaria es menor que la máxima restricción de cardinalidad de una interrelación ternaria entre dos entidades implicadas, entonces las restricciones de la interrelación ternaria son inválidas.*

2.2 Método de Lenzerini y Nobili

El método se basa en las restricciones de cardinalidad, las cuales expresan el número mínimo y máximo de instancias de una interrelación en la cual cada instancia de una entidad está conectada a través de un rol dado y se puede representar de la siguiente manera:

$$E\ (U) \rightarrow (x, y)\ R$$

donde:

- E es una entidad, R una interrelación, U un rol y E está conectada a R a través del rol U;
- "x" es un número positivo, llamado cardinalidad mínima de R con respecto a E a través del rol U;
- "y" es un número positivo o ∞, llamado cardinalidad máxima de R con respecto a E a través del rol U;

- “$y \geq x$”.

Un esquema que no es válido contiene uno o más conjuntos de restricciones de cardinalidad inconsistentes. Para determinar este conjunto de restricciones inconsistentes se usa el siguiente método en el cual se representa el esquema ER mediante un grafo asociado. A continuación se usa la definición de representación de un esquema ER mediante un grafo.

Definición: Dado un esquema S, el grafo asociado G es un multigrafo dirigido (N,A) etiquetado en los arcos donde:

- El conjunto de nodos N corresponde uno a uno al conjunto de clases de S (tipos de entidades y tipos de interrelaciones).
- El conjunto de arcos A es determinado por las siguientes reglas: por cada conexión en S entre una entidad E y una relación R a través del rol U, dos arcos e_1 y e_2 están en A; e_1 está dirigido desde el nodo correspondiente a E hacia el nodo correspondiente a R y es etiquetado con MAX(R,E,U) (cardinalidad máxima); e_2 está dirigido desde el nodo correspondiente a R hacia el nodo correspondiente a E y es etiquetado con 1/MIN(R,E,U), si MIN(R,E,U)≠0 (cardinalidad mínima) o ∞ si MIN(R,E,U)=0.

Un grafo asociado a un diagrama ER es llamado Grafo-ER. La etiqueta de un arco e será denotado por Etiqueta (e). Si Л es una ruta (o ciclo) de un Grafo-ER G, entonces el peso de л (denotado por PESO (л)) es definido como sigue:

$$\text{PESO}(Л) = \prod_{e \in Л} \text{Etiqueta}(e)$$

Si γ es un ciclo y el PESO (γ) < 1 entonces γ es un ciclo crítico.

Una asignación Ø para un Grafo-ER G=(N, A) es mapeado

Ø: N→R+

Asociando a los números racionales positivos con sus nodos.

Se dice que una asociación es correcta si para cada arco e= (n_1, n_2) en A se cumple la siguiente condición:

$$\frac{Ø(n_2)}{Ø(n_1)} \leq \text{Etiqueta}(e)$$

Un Grafo-ER es inconsistente si no existen asociaciones para él y consistente en otro caso.

Se puede observar que, si л es una ruta desde n_a hasta n_b de un Grafo-ER G entonces el PESO (л) representa un umbral para $\frac{Ø(n2)}{Ø(n1)}$ para cada asignación correcta Ø de G:

$$\frac{Ø(n_2)}{Ø(n_1)} \leq \text{PESO}(л) \tag{1}$$

Lema 1: Se tiene que G es un grafo consistente y n_1 y n_2 dos de sus nodos. Tenemos que Φ es la colección de todas las asignaciones correctas para G y П la colección de todas las rutas desde n_2 a n_1 en G. Entonces se cumple que:

$$\max_{\phi \in \Phi} \frac{\phi(n_1)}{\phi(n_2)} = \min_{\pi \in \Pi} \text{PESO}(\pi)$$

Donde, si Л es vacío el lado derecho es ∞.

Prueba: es obvio que el $\max_{\phi \in \Phi} \frac{\phi(n_1)}{\phi(n_2)}$ es menor que $\min_{л \in п}$ PESO(л), entonces se puede considerar sólo el caso donde el lado izquierdo es finito. Tenemos que Ø´ es una correcta asignación tal que $\max_{\phi \in \Phi} \frac{\phi'(n_1)}{\phi'(n_2)}$ es máximo y suponemos que $\max_{\phi \in \Phi} \frac{\phi'(n_1)}{\phi'(n_2)} < \min_{\pi \in \Pi} \text{PESO}(\pi)$. Esto significa que cada ruta $л_i$ en П contiene un arco e_i (desde p hasta q) tal que $\frac{\phi'(n_1)}{\phi'(n_2)} < \textbf{Etiqueta}(e_i)$.

El conjunto de todos los e_i constituye un corte (N_a,N_b) de G, con n_a є N_a y n_b є N_b. Si se multiplica cada Ø´(n) (para n є N_a) por un número positivo se puede obtener una asignación correcta ψ tal que $\frac{\psi(na)}{\psi(nb)} > \frac{Ø'(na)}{Ø'(nb)}$ contradiciendo la hipótesis sobre Ø´.

Teorema 1: *Un Grafo-ER G es inconsistente sí y sólo sí contiene un ciclo crítico.*

Prueba: Si se asume que G contiene un ciclo crítico γ con peso w y suponiendo que Ø es una asignación correcta para G, entonces para cualquier nodo n en γ se tiene que:

$$\frac{Ø(n)}{Ø(n)} \leq w \qquad Ø \qquad (2)$$

Obtenido al aplicar (1) a γ. Desde (2) no puede ser satisfecha por tanto no existe una asignación correcta para G.

Si se asume que G es inconsistente, se tiene que G´ es un subgrafo de máxima consistencia con un arco cualquiera e=<n_a, n_b> de G-G´. El lema 5 encierra que existe al menos una ruta desde n_a hasta n_b en G´ (en otro caso se puede encontrar una asignación correcta para G´U {e} contradiciendo la hipótesis de que G´ es máximamente consistente). Desde el Lema 1 se tiene que Ø es una asignación correcta para G´ y л una ruta desde n_b hasta n_a tal que:

$$\frac{Ø(n_a)}{Ø(n_b)} = PESO(\pi)$$

Si se considera que el ciclo γ está constituido por e y л y el PESO (γ) ≥ 1 entonces

1/<u>PESO</u>(л) ≤ Etiqueta(e). Se tiene que:

$$\frac{Ø(n_a)}{Ø(n_b)} \leq ETIQUETA(e)$$

lo que significa que Ø es correcta para G´U {e}. Si se contradice la hipótesis de que G´ es máximamente consistente entonces se tiene γ es un ciclo crítico.

2.3 Conclusiones parciales

En este capítulo se han expuesto los fundamentos teóricos de dos métodos de validación estructural de esquemas conceptuales ER, los cuales se basan en las restricciones de cardinalidad para determinar si existen o no restricciones de cardinalidad inconsistentes. También se ha expuesto la manera en que ambos métodos determinan las inconsistencias en el esquema conceptual.

3 CASOS DE ESTUDIO

Para la representación de casos de estudio se van a utilizar dos tipos de diagramas ER, uno definido por Elmasri (Elmasri and Navathe 2007) donde se usa "uno" para representar la mínima cardinalidad y "uno" o "muchos" para la máxima, una línea para representar opcionalidad y doble para la obligatoriedad, y el otro por Olivé (Olivé 2007) donde la cardinalidad se define como un par expresado como Card(E_1, E_2,R)=(min, max). Por ejemplo, si se tiene un grupo de Empleados donde cada uno de ellos puede trabajar en un sólo Departamento y en este a la vez trabajan de "uno" a "muchos" Empleados va a quedar representado de la siguiente manera de acuerdo a la notación de ambos autores.

Figura 3.1. Diagrama ER siguiendo la notación de Olivé (min, max).

Figura 3.2. Diagrama ER siguiendo la notación de Elmasri.

A continuación se muestran los casos de estudio a los que se les va a aplicar los dos métodos descritos anteriormente en la literatura. Partiendo de la idea de Hartmann (Hartmann 1998) que plantea que: "el conjunto de restricciones de cardinalidad restrictivo es consistente siempre que el correspondiente conjunto relajado de restricciones de cardinalidad también sea consistente", se puede decir que ambos métodos están descritos para encontrar las inconsistencias de un diagrama ER a través del conjunto de restricciones de cardinalidad asociado al mismo, con la diferencia de que Lenzerini (Lenzerini and Nobili 1990) utilizan el llamado conjunto de restricciones de cardinalidad más restrictivo con valores de 0,1 y mayores que 1 para representar la mínima cardinalidad y 1, mayores que 1 y ∞ para la máxima, mientras que Dullea (Dullea, Song et al. 2003) usan el conjunto de restricciones más relajado con un universo de discurso más amplio usando valores como 1 para representar la mínima cardinalidad y 1 ó M para la máxima.

3.1 Caso de estudio 1

Se quiere representar un grupo de Empleados que pueden dirigir a "muchos" Empleados pero a la vez tienen que ser subordinados por un sólo Director.

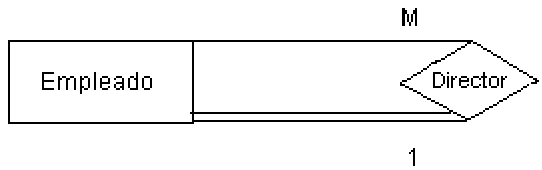


Figura 3.3. Representación del esquema ER según Elmasri y Navathe.

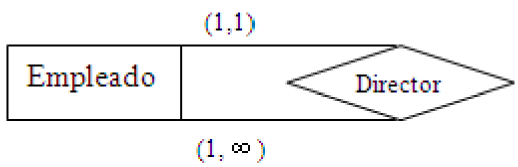
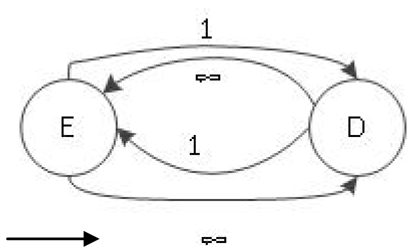


Figura 3.4. Representación del esquema ER según la notación de Olivé.

Método Dullea y Song:

Según el Corolario 3 (Dullea, Song et al. 2003) el esquema no es consistente, lo que significa que no existe un conjunto de instancias válidas que satisfagan el esquema.

Método de Lenzerini y Nobili:



Primera vía

Figura 3.5. Grafo asociado según Lenzerini y Nobili.

Donde E=Empleado y D=Departamento.

PESO (E-D-E)= 1*1=1 $\geq$ 1 Consistente

PESO (E-D-E)= $\infty * \infty = \infty \geq 1$ Consistente

3.2 Caso de estudio 2

Se quiere representar un grupo de Personas relacionadas a través de la interrelación pariente-de donde cada una tiene 2 padres y 3 hijos. El diagrama queda representado de la siguiente manera:

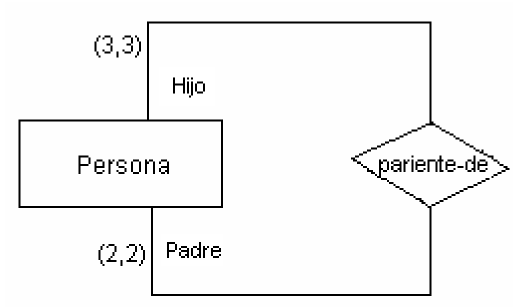


Figura 3.6. Diagrama ER según Olivé

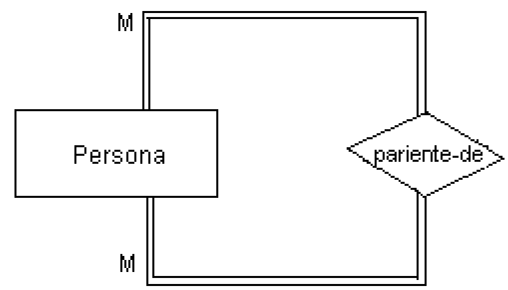


Figura 3.7. Diagrama ER según Elmasri.

Método de Dullea y Song:

Según la regla 4 el diagrama representado anteriormente es una estructura válida.

Método de Lenzerini y Nobili:

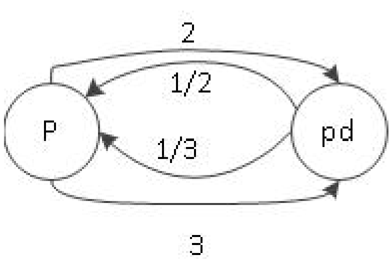


Figura 3.8. Grafo asociado al diagrama ER de la Figura 3.6

Donde P = Persona y p-d = pariente-de.

PESO (P-(p-d) -P) = 3*1/2 = 3/2 ≥ 1

PESO (P-(p-d) -P) = 2*1/3 = 2/3 ≤ 1 Es inconsistente.

De acuerdo al método se puede observar que el diagrama es inconsistente y que existe una contradicción al comparar el resultado de ambos métodos pero si se relaja el conjunto de restricciones de la situación descrita anteriormente se puede observar otro resultado, por ejemplo, si se considera el hecho de que una persona puede tener de ¨uno¨ a ¨tres¨ hijos y que puede tener a la vez de ¨uno¨ a ¨dos¨ padres en el grupo de Personas el grafo asociado queda representado de la siguiente manera:

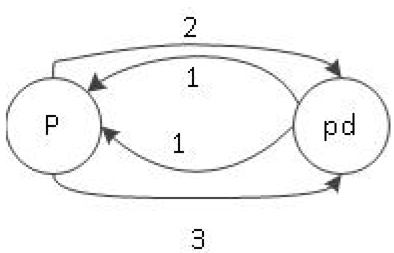


Figura 3.9. Grafo asociado a la figura 3.6 con un conjunto de restricciones más relajadas.

PESO (P-(p-d) -P) = 3*1=3 ≥ 1

PESO (P-(p-d) -P) = 2*1=2 ≥ 1 Consistente

Para este nuevo conjunto de restricciones el diagrama es consistente de acuerdo al método aplicado y en cuanto al método de Dullea y Song se mantiene el resultado anterior porque la interrelación sigue siendo M:M y según la Regla 4 un diagrama ER con esta característica es consistente.

3.3 Caso de estudio 3

En el diagrama de la Figura 3.10 se modelan un grupo de Empleados que pueden trabajar o no en un Departamento y dirigir varios a la vez; un Departamento es dirigido por un único Empleado pero pueden trabajar en el mismo de dos a muchos Empleados.

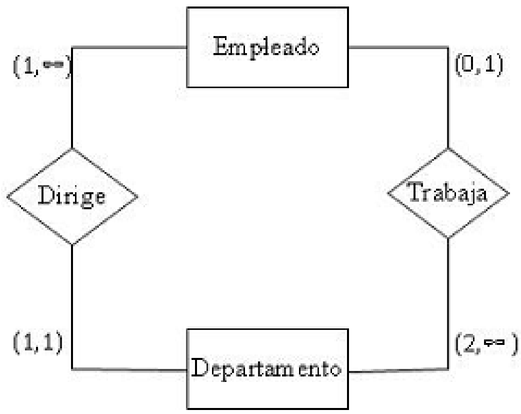


Figura 3.10. Diagrama ER, notación Olivé.

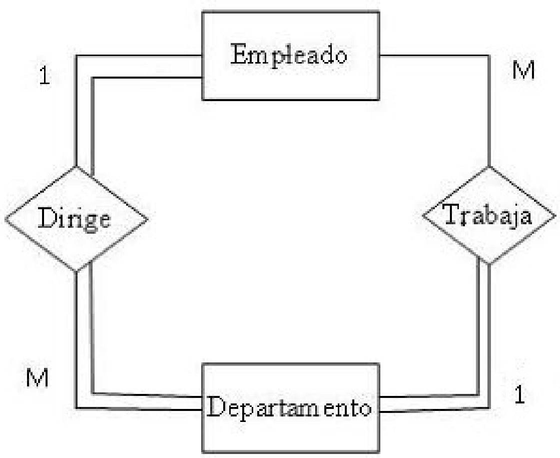


Figura 3.11. Diagrama ER, notación Elmasri y Navathe.

Método de Dullea y Song:

La figura 3.11 muestra un ciclo de interrelaciones binarias, si se analiza la Tabla 2 se deriva que:

B9 (Empleado-Dirige-Departamento): │Empleado│< │Departamento│

B10 (Departamento-Trabaja-Empleado): │ Departamento│< │Empleado│

Se puede observar que existe una contradicción entre B9 y B10 dando como resultado que el esquema es inconsistente, por tanto la figura 3.11 es una relación circular.

Método de Lenzerini y Nobili:

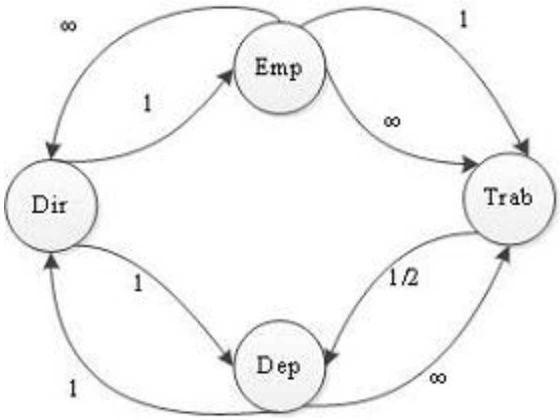


Figura 3.12. Grafo asociado a la Figura 3.10.

PESO (Emp-Dir-Dep-Trab)= ∞*1*∞*∞= ∞ >1

PESO (Emp-Trab-Dep-Dir)= 1*1/2*1*1= 1/2 <1 Es inconsistente

3.4 Caso de estudio 4

Se quiere representar un grupo de fabricantes que compran muchos productos y son atendidos por un único vendedor, los vendedores atienden a muchos fabricantes y venden un tipo de producto en específico por tanto cada tipo de producto es vendido por un único vendedor y comprado por un fabricante. En las figuras siguientes queda modelada la situación descrita.

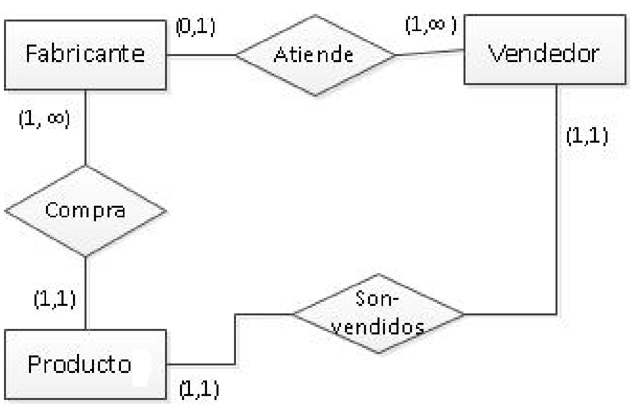


Figura 3.13. Diagrama ER según la notación de Olivé.

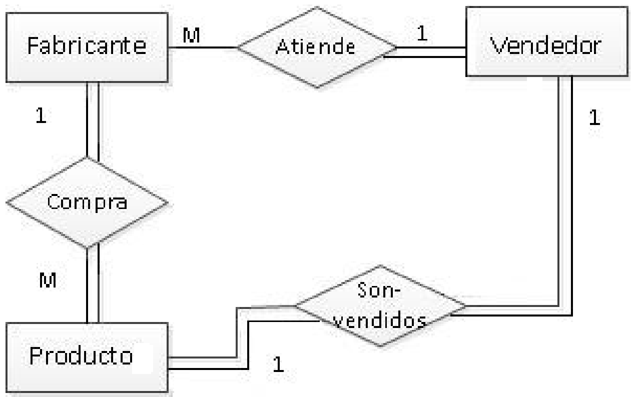


Figura 3.14. Diagrama ER según la notación de Elmasri y Navathe.

Método de Dullea y Song:

De la tabla 2 se deriva que:

B9 (Fabricante-Producto): │ Fabricante │< │Producto│

B1 (Producto-Vendedor): │ Producto │=│ Vendedor │

B10 (Vendedor- Fabricante): │ Vendedor │<│ Fabricante │

Si se sustituye B1 en B10 y se compara con B9 se puede observar que existe inconsistencia entre ambas inecuaciones por lo tanto la estructura es inconsistente y se puede decir que la figura 3.14 según el resultado del método aplicado es una relación circular.

Método de Lenzerini y Nobili:

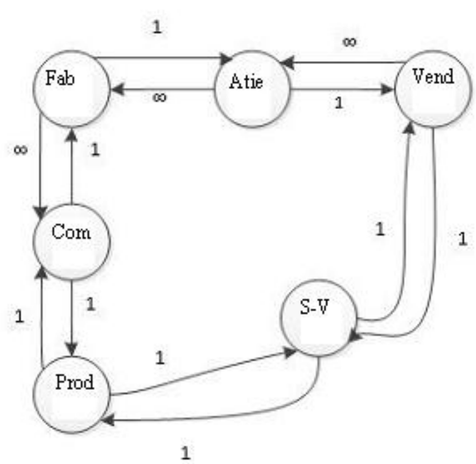


Figura 3.15. Grafo asociado a la figura 3.13.

PESO (Fab-Atie-Vend-(S-V)-Prod-Com- Fab)= $1*1*1*1*1*1= 1 \geq 1$

PESO (Fab-Com-Prod-(S-V)-Vend-Atie-Fab)= $\infty*1*1*1*\infty*\infty= \infty \geq 1$ Consistente

3.5 Caso de estudio 5

Se quiere modelar un grupo de personas que pueden trabajar o ser propietarias de una compañía y se especifica que una compañía es dirigida sólo por una persona y que deben trabajar en la misma como mínimo dos personas.

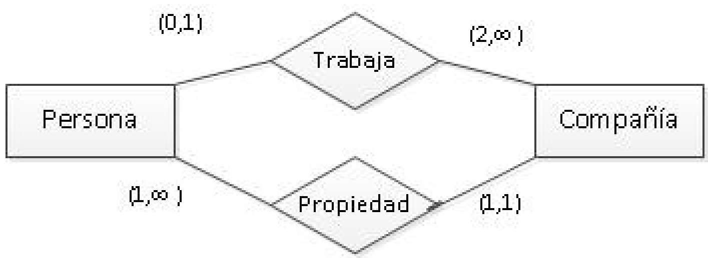


Figura 3.16. Diagrama ER, notación de Olivé.

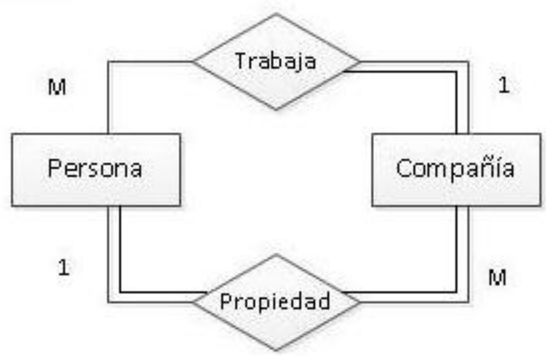


Figura 3.17. Diagrama ER, notación de Elmasri y Navathe.

Método de Dullea y Song:

De acuerdo a la tabla 2 se tiene que:

B9 (Persona-Propiedad-Compañía)= $|\text{Persona}| < |\text{Compañía}|$

B10 (Compañía-Trabaja-Persona)= $|\text{Compañía}| < |\text{Persona}|$

Se puede observar que B9 y B10 se contradicen mutuamente lo que implica que la estructura representada en la figura 3.17 es inconsistente.

Método de Lenzerini y Nobili:

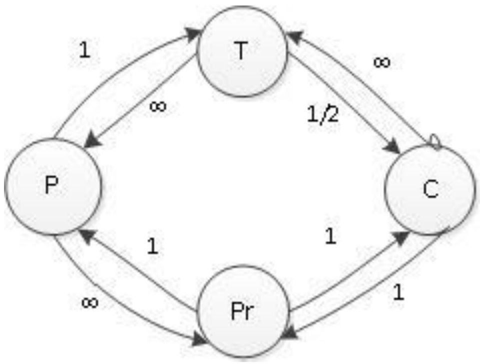


Figura 3.18. Grafo asociado a la figura 3.16.

PESO (T-C-Pr-P-T)= ½*1*1*1= ½ < 1 Inconsistente

PESO (T-P-Pr-C-T)= ∞*∞*1*∞= ∞ ≥ 1

De acuerdo al método el diagrama asociado al grafo es inconsistente porque la multiplicación de uno de sus pesos es < 1.

3.6 Caso de estudio 6

Si se quiere representar una situación dada que consiste en determinado servicio como HMO prestado por un doctor a muchos pacientes, a la vez el doctor ofrece tratamientos a sus pacientes y estos pueden obtener el servicio de HMO en varias ocasiones y este sólo se le realiza a un paciente a la vez. Cada paciente se atiende con un doctor en específico.

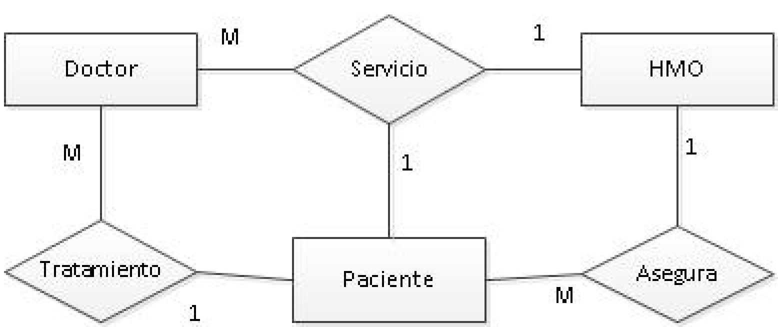


Figura 3.19. Diagrama ER de una interrelación ternaria, notación de Elmasri y Navathe.

Método de Dullea y Song:

De acuerdo con la Regla 13 se puede llegar a la conclusión de que el diagrama es consistente.

Método de Lenzerini y Nobili:

Este método sólo puede ser usado en interrelaciones unarias o binarias.

3.7 Conclusiones parciales

Como resultado de este capítulo se puede concluir que:

1. Ambos métodos se utilizan para determinar la validez de un diagrama ER de acuerdo al conjunto de restricciones de cardinalidad asociadas.
2. El método de Dullea y Song se utiliza en interrelaciones recursivas, binarias y ternarias mientras que el método de Lenzerini y Nobili sólo se puede utilizar con interrelaciones de asociación de hasta grado dos.
3. Al analizar los resultados de los casos de estudio se puede observar que el método de Lenzerini y Nobili no detecta inconsistencias cuando el conjunto de restricciones está formado por 0 y 1 para la cardinalidad mínima y 1 o M para la máxima. En cuanto al método de Dullea y Song se puede decir que presenta dificultades igualmente para detectar las inconsistencias pero sólo cuando el conjunto de restricciones de cardinalidad presenta valores mayores que 1 en la cardinalidad mínima para las interrelaciones recursivas.

CONCLUSIONES

Como resultado de esta tesis, se ha arribado a las conclusiones siguientes:

1. La evaluación del marco teórico mostró que existen varios métodos que permiten determinar la consistencia de un esquema conceptual. Estos métodos tienen en común en que parten de la determinación de los caminos cíclicos dentro del diagrama ER. Algunos de estos métodos difieren en la manera en que determinan la consistencia de las restricciones de cardinalidad, que en este trabajo se han identificado como el método de Lenzerini y Nobili y el método de Dullea y Song.
2. Se formalizaron los algoritmos para el método de Lenzerini y Nobili y el método de Dullea y Song.
3. A través de casos de estudio se comprobó la eficacia de cada uno de estos métodos.
4. Se puede concluir que el método de Dullea y Song puede ser aplicado a una amplia variedad de construcciones del modelo ER y que abarca al otro método en cuanto a las interrelaciones binarias.

RECOMENDACIONES

Derivadas de la investigación realizada, así como de las conclusiones generales emanadas de la misma, se recomienda:

1. Implementar ambos métodos de validación en la nueva versión sobre software libre de la herramienta ERECASE con el propósito de recopilar información sobre la eficacia de ambos métodos.

REFERENCIAS BIBLIOGRÁFICAS

Balaban, M. and A. Maraee (2006). Consistency of UML Class Diagrams with Hierarchy Constraints. NGITS 2006, Springer.

Balaban, M., A. Maraee, et al. (2007). "Reasoning with UML Class Diagrams: Relevance, Problems, and Solutions – a Survey."   Retrieved 2008-09-10, from http://www.cs.bgu.ac.il/~mira/CDReasoning-07.pdf.

Batini, C., S. Ceri, et al. (1992). Conceptual Database Design: An Entity-Relationship Approach. Redwood City, CA., Benjamin/Cummings.

Cadoli, M., D. Calvanese, et al. (2004). Finite satisfiability of UML class diagrams by constraint programming. International Workshop on Description Logics (DL'2004).

Calì, A. (2007). Querying Incomplete Data with Logic Programs: ER Strikes Back. ER 2008, Auckland, New Zealand, Springer.

Calvanese, D. and M. Lenzerini (1994). On the Interaction Between ISA and Cardinality Constraints. Tenth International Conference on Data Engineering, Houston, Texas, USA, IEEE Computer Society Press.

Combi, C., S. Degani, et al. (2008). Capturing Temporal Constraints in Temporal ER Models. ER 2008, Springer.

Czejdo, B., R. Elmasri, et al. (1990). "A graphical data manipulation language for an extended entity-relationship model." IEEE Computer 23(3): 26-35.

Chen, P. (1976). "The entity-relationship model: Toward a unified view of data." ACM Transactions on Database Systems 1(1): 9-36.

Chen, P. P. (2006). Suggested Research Directions for a New Frontier – Active Conceptual Modeling. ER 2006, Springer.

Date, C. J. (2003). An Introduction to Database Systems. Boston, MA, Addison-Wesley.

Dullea, J. and I.-Y. Song (1997). An Analysis of Cardinality Constraints in Redundant Relationships. 6th InternationalConference on Information and Knowledge Management (CIKM '97), Las Vegas, Nevada, USA.

Dullea, J. and I.-Y. Song (1998). An Analysis of Structural Validity of Ternary Relationships in Entity Relationship Modeling. 7th International Conference on Information and Knowledge Management (CIKM ‘98). Washington, D.C.

Dullea, J. and I.-Y. Song (1998). An Analysis of the Structural Validity of Unary and Binary Relationships in Entity Relationship Modeling. 4th International Conference on Computer Science and Informatics., Research Triangle Park, NC, USA.

Dullea, J. and I.-Y. Song (1999). A Taxonomy of Recursive Relationships and Their Structural Validity in ER Modeling. ER 1999, Paris, France, Springer.

Dullea, J., I.-Y. Song, et al. (2003). "An Analysis of Structural Validity in Entity-Relationship Modeling." Data & Knowledge Engineering 47: 167-205.

Elmasri, R. and S. B. Navathe (2007). Fundamentals of Database Systems. Boston, USA, Addison-Wesley.

Elmasri, R., J. Weeldreyer, et al. (1985). "The Category Concept: An Extension to the Entity-Relationship Model." Data & Knowledge Engineering 1(1): 75-116.

Garcia-Molina, H., J. D. Ullman, et al. (2008). Database Systems. The Complete Book, Prentice Hall.

García, L. and M. Montes de Oca (2005). Sistemas de Bases de Datos: Modelación y Diseño. La Habana, Editorial Felix Varela.

Hartmann, S. (1998). On the Consistency of Int-cardinality Constraints. ER 1998, Singapore, Singapore.

Hartmann, S. (2000). On Interactions of Cardinality Constraints, Keys and Functional Dependencies. FoIKS 2000, Burg, Germany, Springer.

Hartmann, S. (2001). Coping with Inconsistent Constraint Specifications. ER 2001, Yokohama, Japan, Springer.

Hartmann, S. (2001). "Decomposing relationship types by pivoting and schema equivalence." Data & Knowledge Engineering(39): 75-99.

Hartmann, S. (2001). "On the implication problem for cardinality constraints and functional dependencies." Annals of Mathematics and Artificial Intelligence 33(2-4): 253-307.

Hartmann, S. (2003). Soft Constraints and Heuristic Constraint Correction in Entity-Relationship Modelling. International workshop on semantics in databases, Dagstuhl Castle, Germany, Springer.

Hartmann, S., S. Link, et al. (2009). "Constraint acquisition for Entity-Relationship models." Data & Knowledge Engineering 68(10): 1128-1155.

Hay, D. C. (1999). "A Comparison of Data Modeling Techniques." Retrieved 2007-12-21, from http://www.essentialstrategies.com/publications/modeling/compare.htm.

Hohenstein, U. and M. Gogolla (1988). A calculus for an extended entity-relationship model incorporating arbitrary data operations and aggegate funtions. ER 1988, North-Holland Publishing Co.

Jones, T. H. and I.-Y. Song (1996). "Analysis of Binary/Ternary Cardinality Combinations in Entity-Relationship Modeling." Data & Knowledge Engineering 19(1): 39-64.

Jones, T. H. and I.-Y. Song (2000). "Binary Equivalents of Ternary Relationships in Entity-Relationship Modeling: a Logical Decomposition Approach." Journal of Database Management(April-June): 12-19.

Jones, T. H. and I.-Y. Song (2002). Ternary Relationships: Semantic Requirements and Logically Correct Alternatives. Advanced topics in database research. Hershey, PA, USA, IGI Publishing. 1: 17-33.

Kolapalli, D. (2008). AN ANALYSIS OF STRUCTURAL AND SEMANTIC VALIDITY IN THE DATA MODEL. Department of Computers science Villanova University. Master of science in computers science.

Lenzerini, M. and P. Nobili (1990). "On the satisfiability of dependency constraints in entity–relationship schemata." Information Systems 15(4): 453-461.

Lenzerini, M. and G. Santucci (1983). Cardinality constraints in the entity-relationship model. ER 1983, North-Holland.

Markowitz, V. and A. Shoshani (1992). "Representing extended entity-relationship structures in relational databases: A modular approach." ACM Transactions On Database Systems 17(3): 423-464.

McAllister, A. (1998). "Complete rules for n-ary relationship cardinality constraints." Data & Knowledge Engineering 27: 255-288.

Murthy, S., L. Delcambre, et al. (2006). Explicitly Representing Superimposed Information in a Conceptual Model. ER 2008, Tucson, AZ, USA, Springer.

Olivé, A. (2007). Cardinalitiy Constraints. Conceptual Modeling of Information Systems, Springer: 83-102.

Olivé, A. (2007). Integrity Constraints. Conceptual Modeling of Information Systems, Springer: 83-102.

Piattini, M. G. and O. Díaz (2000). Advanced Database Technology and Design, Artech House Inc.

Ponniah, P. (2003). Database Design and Development, Wiley-Interscience.

Santos, I.-J., P. Martínez, et al. (2007). On the semi-automatic validation and decomposition of ternary relationships with optional elements. International Conference on Enterprise Information Systems, Funchal, Madeira - Portugal.

Shanks, G., E. Tansley, et al. (2003). "Using Ontology to Validate Conceptual Models." Communications of the ACM 46(10): 85-89.

Silberschatz, A., H. Korth, et al. (2006). Sistemas de Bases de Datos, Pearson Addison-Wesley.

Song, I.-Y., M. Evans, et al. (1995). "A comparative analysis of entity-relationship diagrams." Journal of Computer and Software Engennering 3(4): 427-459.

Song, I.-Y. and T. H. Jones (1993). An analysis of binary relationships within ternary relationships. ER 1993, Arlington, Texas, USA, Springer.

Teorey, T., D. Yang, et al. (1986). "A logical design methodology for relational databases using the extended E-R model." ACM Computing Surveys 18(2): 197-222.

Teorey, T. J., S. S. Lightstone, et al. (2006). Database Modeling and Design: Logical Design, Morgan Kaufmann.

Thalheim, B. (1992). Fundamentals of cardinality constraints. ER 1992, Karlsruhe, Germany, Springer.

Thalheim, B. (1999). The Strength of ER Modeling. Current Issues in Conceptual Modelling, Springer.

Thalheim, B. (2000). Integrity Constraints. Entity-Relationship Modeling. Foundations of Database Technology, Springer-Verlag: 105-218.

Ullman, J. D. and J. Widom (2007). First Course in Database Systems, Prentice Hall.