

FORMAS NORMALES MODELO ENTIDAD RELACION.

Alonso Márquez y Maria Esther Vidal.

Universidad Simón Bolívar.

Caracas-Venezuela.

RESUMEN

Uno de los Modelos de Datos más ampliamente usado en la etapa de diseño conceptual debido a su legibilidad, expresividad y economía conceptual, es el Modelo Entidad Relación.

Pese a esto, no existe una definición formal satisfactoria que permita interpretar de una manera no ambigua un esquema E-R. Tampoco existe una Teoría de Normalización que permita decidir si un esquema E-R es bueno, es decir, está libre de las anomalías de inserción, modificación y eliminación.

En este artículo, se presenta una especificación formal del Modelo E-R a partir de la cual se definen formas normales para esquemas E-R. Los esquemas E-R que cumplen estas formas normales tienen la propiedad que el esquema equivalente en Relacional, cumple como mínimo 3FN.

Además, se presentan técnicas de traducción de esquemas E-R a esquemas E-R normalizados y técnicas de traducción de esquemas E-R a esquemas Relacionales + Restricciones de Integridad que representan toda la semántica del esquema E-R.

1. INTRODUCCION.

Una base de datos es una representación de los datos operativos de una organización[8].

Tradicionalmente en el área de Base de Datos se denomina Modelo de Datos a la herramienta de modelaje de datos o sistema formal de especificación que permite representar las propiedades y características de los datos operativos de una organización, es decir, es la herramienta utilizada para definir la base de datos[19].

El resultado de la utilización de un modelo de datos para especificar una situación en particular se denomina esquema[19].

Dado un modelo de datos y una situación particular, existen muchos esquemas que representan dicha situación. Dependiendo de que tanto se acerque a la realidad, se puede clasificar el esquema en bueno o malo.

Luego sería de gran interés, conocer las características de los buenos esquemas de un modelo de datos particular y técnicas para transformar esquemas malos en buenos.

La teoría que engloba todas estas técnicas se denomina Teoría de Normalización, donde el nombre viene de suponer que un esquema se considera "bueno" si cumple una cierta forma normal.

Para el único modelo de datos que ha sido definida una Teoría de Normalización es para el modelo Relacional[5,6,20]. Este modelo debido a las abstracciones que ofrece, no es utilizado como herramienta de especificación conceptual. Por el contrario, el modelo de datos más ampliamente usado es el modelo Entidad Relación[7,12,16]. A pesar de esto, no existe definida una teoría de normalización para el mismo.

Por esta razón, se está realizando un trabajo que tiene como objetivos fundamentales dar una definición formal del modelo E-R

y sobre ésta, definir una Teoría de Normalización.

En este artículo serán presentados algunos de los logros alcanzados en este trabajo: elementos de la especificación formal del modelo, formas normales de esquemas E-R, técnicas de normalización y técnicas de traducción a esquemas relacionales.

Para dar la especificación del modelo Entidad Relación se utilizará el Lenguaje de la Lógica de Predicados y conceptos de la Teoría de Conjuntos. Las formas normales definidas poseen la propiedad que el esquema equivalente en relacional cumple como mínimo la 3FN.

Las técnicas de normalización de esquemas E-R presentadas cumplen la propiedad que en el esquema resultante se mantiene toda la información.

Por otro lado, las técnicas de traducción cumplen la propiedad que toda la información representada en el esquema E-R también se representa en el Relacional, ya que aquellas abstracciones que no tienen abstracción equivalente son representadas a través de restricciones de integridad.

Para presentar estos logros el resto de este artículo se estructura de la siguiente manera: en la sección II se muestran los antecedentes del problema presentado. En la sección III se da una especificación del Modelo E-R. En la sección IV se definen las Formas Normales de un esquema E-R. En la sección V se presentan técnicas de normalización. En la sección VI se presentan técnicas de traducción a Relacional. Finalmente, en la sección VII se presentan las conclusiones y posibles extensiones.

2. ANTECEDENTES.

Son bien conocidas las anomalías de inserción, modificación y eliminación que se pueden presentar en una Base de Datos como consecuencia de no realizar un buen diseño de la misma[6,20]. Para evitar estos problemas se pueden adoptar alguna de las siguientes soluciones:

a) Utilizar un modelo de datos que permita representar restricciones de integridad, es decir, representar cuales son los estados de la base de datos que no son permitidos.

b) Diseñar esquemas que cumplan ciertas características, es decir, que satisfagan una cierta forma normal.

La solución más ampliamente usada y estudiada es la de llevar el esquema hasta una cierta forma normal.

En esta dirección, se ha definido una Teoría de Normalización para el modelo de datos Relacional.

El modelo de datos Relacional[5,6,8,20], es un modelo que ofrece la abstracción de relación, siendo una relación un subconjunto del producto cartesiano de los dominios sobre los cuales la misma está definida. Una relación puede ser vista como una tabla donde las columnas representan los atributos de la relación y, las filas o tuplas, las instancias u ocurrencias de ésta.

Las restricciones que se permiten representar a través del modelo son de dependencia funcional y multivaluada entre atributos y las inherentes a éste, identidad e integridad referencial.

Dado que ofrece una única abstracción, el modelo presenta las siguientes características: economía conceptual, sobre carga

de abstracciones y poca expresividad. Un ejemplo donde se muestran las limitaciones expresivas del mismo es el siguiente:

Representar a través de un esquema relacional las materias que ha inscrito en un trimestre un estudiante cumpliendo la restricción que como mínimo puede inscribir 2 materias y como máximo 5.

Las relaciones EST y MAT mantienen la información de los estudiantes y de las materias respectivamente.

EST(CARNET,NOMBRE) CLAVE:CARNET MAT(COD-MAT,N-MAT) CLAVE:COD-MAT

Por otro lado, la relación INSCRIBE mantiene la información de cada estudiante con las materias que inscribe.

INSCRIBE(COD-MAT,CARNET) CLAVE: COD-MAT,CARNET.

CLAVES FORANEAS: COD-MAT NOT NULL; CARNET NOT NULL.

Pero, la restricción de que dado un estudiante puede tener asociado como mínimo dos materias y como máximo cinco, no se puede representar a través de las estructuras del modelo.

Por causa de su poca expresividad, el modelo Relacional no se utiliza como herramienta de especificación en la etapa de diseño conceptual. Por el contrario, para esta etapa se utilizan modelos de datos que permiten representar de una manera más fiel la situación a modelar. Por esta razón, se utilizan los modelos de datos semánticos, los cuales son herramientas de especificación que permiten representar gran parte de las abstracciones que existen en el mundo real. Entre los modelos semánticos, el más utilizado es el modelo Entidad Relación.

A causa de su economía conceptual, el modelo Relacional es el modelo de datos que más manejadores comerciales ofrecen, es

decir, es el modelo que más se utiliza en la etapa de implementación.

Como se dijo anteriormente, existe una Teoría de Normalización para este modelo. En ésta se han definido un grupo de formas normales[5,8,10,20], algoritmos de normalización[0,2,4,9] y criterios de equivalencia de esquemas[20], entre otras cosas.

Las formas normales definidas son seis de las cuales las cuatro primeras imponen restricciones sobre las dependencias funcionales que se pueden cumplir en una relación. De éstas son de interés la 3FN y BCNF, ya que la 2FN es un caso particular de la 3FN y la 1FN impone que las relaciones sean planas.

Considerando 3FN y BCNF lo más razonable de alcanzar es 3FN ya que, existen esquemas relacionales que no pueden alcanzar BCNF sin que exista pérdida de información[8,20]. Además se, ha demostrado que determinar si un esquema relacional tiene un esquema equivalente en BCNF es NP-Completo[3,14].

Dado que solo existe definida una teoría de normalización para el modelo relacional y debido a que éste no se utiliza en la etapa de diseño conceptual, para poder diseñar un esquema normalizado se deben los siguientes pasos:

DISENO ESQUEMA E-R

TRADUCCION

⋮
V

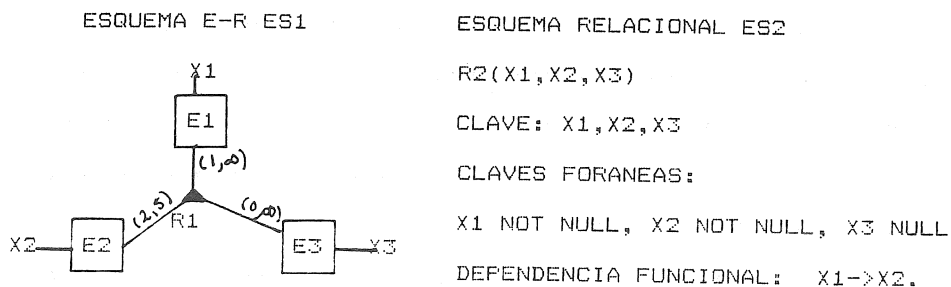
NORMALIZACION

ESQUEMA RELACIONAL----->ESQUEMA RELACIONAL
NORMALIZADO.

Posponiendo así el proceso de normalización para la etapa de diseño de la implementación, impidiendo de esta manera, en muchos casos, que el diseñador del diseño conceptual pueda verificar si

el esquema normalizado representa lo que realmente se quería representar.

Además de esta limitación, el proceso de normalización a nivel relacional puede presentar problemas. Considérese por ejemplo la siguiente situación:



En la traducción del esquema ES1 al esquema ES2 ha habido pérdida de información ya que no se ha podido representar la cardinalidad de E2 en R1.

Siguiendo la definición de 3FN, ES2 no está normalizado. El esquema ES3 es el esquema equivalente normalizado:

ESQUEMA RELACIONAL ES3

$R21(X1, X2)$ CLAVE: $X1$; $R22(X1, X3)$ CLAVE: $X1, X3$.

En el proceso de traducción de ES2 a ES3 ha habido pérdida de información ya que, sobre los atributos $X1$, $X2$ y $X3$ de la relación R1 existían restricciones de integridad referencial que en los atributos de R21 y R22 no existen. Esto es debido a que el proceso de normalización se normalizan las estructuras, pero no, las restricciones de integridad. Apartando problemas como éste, ¿qué ocurriría si en lugar de utilizar el modelo de datos Relacional se utiliza otro modelo de datos para implementar la base de datos?. ¿Es que las anomalías solo se presentan si no se

hace un buen diseño de esquemas relacionales?.

La respuesta es no. En todo esquema que no esté bien diseñado se presentan estos problemas, luego aunque no se implemente con el modelo relacional, se debería poder garantizar la bondad de la base de datos a implementar.

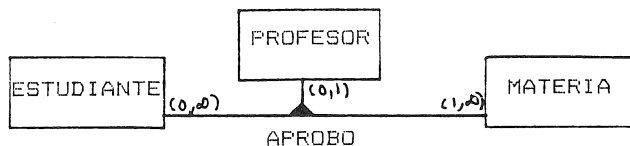
De esta forma, más que tener la teoría de normalización para el modelo Relacional, es necesario tener definida una teoría para un modelo que pueda ser utilizado en la etapa de diseño conceptual, para así garantizar que el esquema conceptual está libre de todos estos problemas.

Como se dijo anteriormente, el modelo E-R es el más ampliamente usado como herramienta de especificación conceptual, luego una posible solución es definir una teoría de normalización para este modelo.

3. MODELO ENTIDAD RELACION.

El modelo Entidad-Relación fue definido de una manera informal en 1976 por Peter Chen[7]. A pesar de su gran aceptabilidad entre los diseñadores de bases de datos y aunque ha sido ampliamente estudiado y analizado en la literatura[15,17,18], no existe una definición formal satisfactoria que permita interpretar un esquema Entidad Relación sin ambigüedad, lo cual trae como consecuencia la existencia de esquemas que pueden ser interpretados de diferentes maneras.

Por ejemplo considérese el siguiente esquema E-R:



A partir de la definición dada por Chen, el límite $(0,1)$ de

la cardinalidad del conjunto entidad PROFESOR en APROBO, podría interpretarse de dos maneras:

a) Como una función parcial f_1 tal que,

$f_1: \text{ESTUDIANTE} \times \text{MATERIA} \rightarrow \text{PROFESOR}$.

Así cada (est, mat) para el cual no esté definido f_1 puede cumplir cualquiera de las siguientes condiciones:

a1) f_1 no está definido para (est, mat) porque est no aprobó mat , razón por la cual, el profesor con el cual est aprobó mat es indeterminado.

a2) f_1 no está definido para (est, mat) porque est aprobó mat por libre escolaridad, razón por la cual, no existe en PROFESOR un profesor con el cual haya aprobado la materia.

Las desventajas de interpretar la cardinalidad de esta forma es que en primer lugar no hay forma de diferenciar cual de las situaciones se está representando y, en segundo lugar, a menos que todos los elementos del producto cartesiano tengan sentido dentro de la relación, el límite inferior será cero.

b) Como una función f_1 total tal que,

$f_1: \text{EST_MAT} \rightarrow \text{PROFESOR} \cup \{\text{NULO}\}$, donde

$\text{EST_MAT} = \{(\text{est}, \text{mat}) / \text{est} \in \text{ESTUDIANTE}, \text{mat} \in \text{MATERIA} \text{ y } \text{est aprobó mat}\}$

Siguiendo esta interpretación, el límite inferior cero se interpreta como que existen elementos (est, mat) que a pesar de cumplir que " est aprobó mat " no tienen asociado un elemento en PROFESOR.

De esta forma, los límites de la cardinalidad de un conjunto de objetos O_n en un conjunto relación R que relaciona los conjuntos de objetos $O_1, \dots, O_n$ se interpretan como el número

mínimo y máximo de ocurrencias de O_n con los cuales se pueden relacionar una ocurrencia de $O_1 \times \dots \times O_{n-1}$ que tenga sentido dentro de R .

A lo largo de este artículo se seguirá la interpretación "b" de la cardinalidad.

Ambigüedades como ésta crean la necesidad de dar una definición formal del modelo, la cual pueda ser utilizada como base en la definición de la teoría de normalización.

Para dar esta especificación se ha utilizado el Lenguaje de la Lógica de Predicados y conceptos de la Teoría de Conjuntos. La misma está dividida en un conjunto de axiomas mediante los cuales se define el significado y las equivalencias de cada una de las abstracciones y, un conjunto de restricciones a través de las cuales, se presentan las condiciones que deben cumplirse cada vez que una abstracción es definida.

Así si se denota con EC_i el esquema por comprensión de BD_i , $AXIOMAS(EC_i)$ el conjunto de las fórmulas lógicas que expresan el significado de cada una de las estructuras de EC_i , $EE(EC_i)$ el esquema por extensión de EC_i , RI las restricciones de integridad entonces

$\forall EC_i \text{ } AXIOMAS(EC_i) \cup EE(EC_i) \cup EC_i \models RI$, es decir, todas las estructuras de interpretación que sean modelos del conjunto de fórmulas $AXIOMAS(EC_i) \cup EE(EC_i) \cup EC_i$ son modelos también del conjunto de restricciones de integridad [1].

El modelo E-R utilizado es una modificación del modelo original definido por Chen al cual para aumentar su poder expresivo se han agregado: conjuntos de relaciones que relacionan

conjuntos de entidades y/o conjuntos de relaciones. Dependencias funcionales entre los atributos de un objeto. Límite inferior en la cardinalidad con la que participa un conjunto objeto en un conjunto relación.

Las abstracciones ofrecidas por el modelo E-R a utilizar son:

Conjunto de Entidades: grupo de entes que tienen propiedades en común. En la axiomatización se denotará un conjunto entidad cuyo nombre es E a través del siguiente predicado: nombre-conjunto-entidad(E).

Entidad: instancia u ocurrencia de un conjunto entidad. Si E es el nombre de un conjunto entidad y e es una instancia de E entonces entidad(e, E).

Conjunto Relación: asociación entre conjuntos de entidades o conjuntos de relaciones. Si R es el nombre del conjunto relación que relaciona a los conjuntos de entidades o de relaciones con nombres $O_1, \dots, O_n$, entonces nombre-conjunto-relacion($R, O_1, \dots, O_n$).

Relación: instancia u ocurrencia de un conjunto relación. Si $(o_1, \dots, o_n)$ es una instancia del conjunto relación R que relaciona los conjuntos entidad o relación $O_1, \dots, O_n$ de los cuales $o_1, \dots, o_n$ son instancias respectivamente, entonces se cumple el siguiente predicado: relacion($R(o_1, \dots, o_n)$).

Conjunto Valor: conjunto de valores que puede tomar una característica de un conjunto entidad o relación. Si V es el nombre de un conjunto valor entonces nombre-conjunto-valor(V).

Valor: instancia u ocurrencia de un conjunto valor. Si v es una instancia del conjunto valor V entonces $V(v)$.

Atributo: es una característica de un conjunto entidad o

relación. Si *Atr* es un atributo del conjunto entidad o relación *O* que puede tomar valores del conjunto de valores *V* entonces atribucion(*Atr*,*O*,*V*).

Objeto: entidad o relación. Si *o* es una instancia del conjunto entidad o relación *O* entonces objeto(*o*,*O*).

Conjunto Objeto: conjunto entidad o relación. Si *O* es un conjunto entidad o relación entonces nombre-conjunto-objeto(*O*).

Todas estas abstracciones cumplen los siguientes axiomas:

objeto(*e*,*E*) <= entidad(*e*,*E*).

objeto(*R*(*o1*,...,*on*),*R*) <= relacion(*R*(*o1*,...,*on*)).

nombre-conjunto-objeto(*E*) <= nombre-conjunto-entidad(*E*).

nombre-conjunto-objeto(*R*) <= nombre-conjunto-objeto(*O1*) &...&
 nombre-conjunto-objeto(*On*) &
 nombre-conjunto-relacion(*R*,*O1*,...,*On*)

nombre-conjunto-entidad(*E*) <= entidad(*e*,*E*).

nombre-conjunto-relacion(*R*,*O1*,...,*On*) <= objeto(*o1*,*O1*)&...&
 objeto(*on*,*On*) &
 relacion(*R*(*o1*,...,*on*)).

nombre-conjunto-valor(*V*) <= *V*=entero v *V*=natural v *V*=caracter v
V=cad-dig v *V*=cad-car.

nombre-conjunto-valor(*V1*x...x*Vn*) <= nombre-conjunto-valor(*V1*)&...&
 nombre-conjunto-valor(*Vn*).

Por ejemplo cada vez que se defina *At* como un atributo de un conjunto de objeto *O* sobre el conjunto de valores *V* se debe satisfacer la siguiente restricción de integridad:

VoVoVAtVv(nombre-conjunto-objeto(*O*) & objeto(*o*,*O*) &
 nombre-conjunto-valor(*V*) & atribucion(*At*,*O*,*V*)=>
 V(*At*(*o*)) v *At*(*o*)=NULO).

Por otro lado, además de estas abstracciones, el modelo permite representar las siguientes restricciones:

De cardinalidad de un conjunto de objetos *O* en un conjunto relación *R*. Así si *On* es el nombre de un conjunto de objetos, *R* /

es el nombre de un conjunto relación que relaciona a los conjuntos de objetos $O_1, \dots, O_n$ y la cardinalidad de O_n en R es $\min, \max$ entonces $\text{cardinalidad}(R, O_n, (\min, \max))$.

De dependencia funcional entre los atributos de un conjunto de objetos. Si los atributos $At_1, \dots, At_n$ implican funcionalmente al atributo At en el conjunto de objetos O entonces $\text{dep-fun}((At_1, \dots, At_n), At, O)$.

Luego cada vez que $\text{cardinalidad}(R, O_n, (\min, \max))$ se cumpla, se debe satisfacer la siguiente restricción de integridad:

Si $o_1, \dots, o_{n-1}$ son elementos del dominio de O_n en R , es decir, pertenece al conjunto de elementos de $O_1 \times \dots \times O_{n-1}$ que tienen sentido dentro de R , entonces $o_1, \dots, o_{n-1}$ se relaciona a través de R como mínimo con $\min$ elementos de O_n y como máximo con $\max$.

$\text{cardinalidad}(R, O_n, (\min, \max)) \leq \Rightarrow$

$\forall o_1 \dots \forall o_{n-1} (\text{objeto}(o_1, O_1) \ \& \dots \ \& \ \text{objeto}(o_{n-1}, O_{n-1}) \ \& \ \text{nombre-dominio}(R_1, R, O_n) \ \& \ \text{relacion}(R_1(o_1, \dots, o_{n-1})) \Rightarrow \min = \< \{ (o_1, \dots, o_n) \mid \text{relacion}(R(o_1, \dots, o_n)) \} \mid \leq \max)$

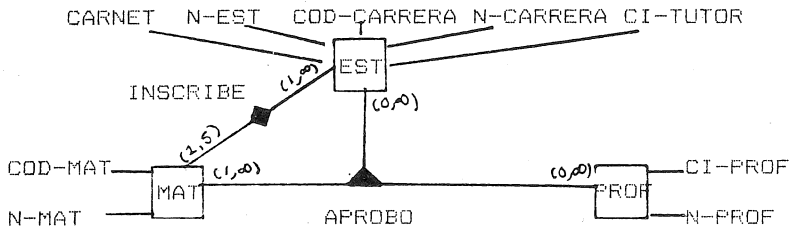
El dominio de un conjunto de objetos O_n en el conjunto relación R es R_1 si, R_1 es el conjunto de todos los elementos de $O_1 \times \dots \times O_{n-1}$ que están relacionados con algún elemento de O_n o con el valor NULO. El valor NULO se interpreta como un elemento que no pertenece a O_n y es diferente al indeterminado y al desconocido.

$\text{nombre-dominio}(R_1, R, O_n) \leq \Rightarrow$

$\forall o_1 \dots \forall o_{n-1} (\text{relacion}(R_1(o_1, \dots, o_{n-1})) \Rightarrow \exists o_n (\text{objeto}(o_n, O_n) \ \& \ \text{relacion}(R(o_1, \dots, o_n))) \vee (\text{nombre-conjunto-relacion}(R, O_1, \dots, O_n \cup \{NULO\}) \Rightarrow \text{relacion}(R(o_1, \dots, NULO))))$.

4. FORMAS NORMALES MODELO ENTIDAD RELACION

Considérese el siguiente esquema E-R que representa la información de un estudiante, la carrera que cursa, las materias que inscribe, las materias que ha aprobado con un profesor y el profesor que es su tutor.



En el conjunto entidad EST, CARNET es la clave, además, existe la dependencia funcional COD-CARRERA $\rightarrow$ N-CARRERA.

En el conjunto relación APROBO existe la dependencia funcional EST, MAT $\rightarrow$ PROF.

Considérense las siguientes operaciones sobre el esquema:

a) Modificar el nombre de la carrera Física Aplicada por Ingeniería Física. Para mantener la consistencia, se deberían actualizar todas las entidades de todos los estudiantes que cursen Física Aplica. Si queda alguno sin actualizar, se genera una inconsistencia.

b) Agregar la carrera Ingeniería Geofísica la cual no estudia algún estudiante. Dado que no puede existir un estudiante sin valor en su CARNET, no se puede agregar mientras no exista un estudiante en la CARRERA.

c) Eliminar el único estudiante que pertenecía a la carrera Ingeniería de Suelos. Originando con esto que se pierda la información de la carrera.

d) Agregar al estudiante Luis Perez el tutor cuya CI-TUTOR

es 5313888. Qué ocurre si el tutor no pertenece al staff de profesores?. Se está creando una inconsistencia.

e) Relacionar a través de APROBO al estudiante est1 con la materia mat1 y el profesor prof1 y, al estudiante est1 con la materia mat2 y el profesor prof2. Dado que (est,mat) se relaciona a través de APROBO con prof1 y prof2, se viola la dependencia funcional EST,MAT-->PROF.

Estos problemas se presentan en el esquema porque en el conjunto entidad EST se mantienen dependencias funcionales que no se expresan explícitamente y se hace referencia a los elementos de un conjunto entidad PROF y no se establece explícitamente la restricción de integridad referencial. Además entre los elementos del conjunto relación APROBO se mantiene una dependencia funcional que tampoco se expresa explícitamente.

Para evitar estos tipos de problemas se definen las siguientes formas normales para el modelo E-R.

Un esquema E-R está normalizado si todo conjunto entidad y conjunto relación del esquema está normalizado.

nombre-esquema(E1) & normalizado(E1) <=>

VEVR(nombre-conjunto-entidad(E) & nombre-conjunto-relacion(R) & conjunto-entidad(E,E1) & conjunto-relacion(R,E1) => normalizado(E) & normalizado(R)).

Sea E un conjunto entidad se dice que E está normalizado si todo atributo depende únicamente de las claves de E y no existe un atributo Atr en E tal que exista otra entidad tal que At sea clave o parte de la clave.

nombre-conjunto-entidad(E) & normalizado(E) $\Leftrightarrow$

$$\forall la \forall At (clave(la, E) \& atribucion(At, E, V) \Rightarrow dep-fun(la, At, E) \& \neg \exists la1 (\neg clave(la1, E) \& dep-fun(la1, At, E))) \& \neg \exists Atr (atribucion(Atr, E, V) \& \exists E1 (nombre-conjunto-entidad(E1) \& clave((At1, \dots, Atr), E1))).$$

Cumpliendo esta forma normal, el conjunto entidad se debería descomponer en dos conjuntos entidad EST y CARRERA y, en dos conjuntos relación ES-TUTOR y ESTUDIA. El conjunto entidad EST con atributos CARNET, N-EST. El conjunto entidad CARRERA con atributos COD-CARRERA, N-CARRERA. El conjunto relación ES-TUTOR relaciona EST con PROF y, cardinalidad(ES-TUTOR, EST, $(0, \infty)$) y cardinalidad(ES-TUTOR, PROF, $(0, 1)$). El conjunto relación ESTUDIA relaciona EST con CARRERA y, cardinalidad(ESTUDIA, CARRERA, $(0, 1)$) y cardinalidad(ESTUDIA, EST, $(0, \infty)$).

En esta descomposición los problemas que se presentaban anteriormente en el conjunto entidad EST ya no se presentan.

Falta ahora solucionar los problemas que se presentan en el conjunto entidad APROBO.

Sea R un conjunto relación, se dice que R está normalizado si no existe un subconjunto S de la clave tal que exista una dependencia funcional X que se mantenga en R y S sea implicante de X, todo elemento que no pertenezca a una clave o sea atributo de R debe depender únicamente de las claves de R y, no existe un atributo Atr en R tal que exista otra entidad tal que At sea clave o parte de la clave.

nombre-conjunto-relacion(R, O1, ..., On) & normalizado(R) $\Leftrightarrow$

$$\forall la (clave(la, R) \& \neg \exists la1 (sub-con-propio(la1, la) \& implicante(la1, R))) \& \forall at (elemento(at, R) \Rightarrow \neg \exists la (\neg clave(la, R) \& de-fun((la, at), R))) \& \neg \exists Atr (atribucion(Atr, R, V) \& \exists E1 (nombre-conjunto-entidad(E1) \& clave((At1, \dots, Atr), E1))).$$

Sea X una dependencia funcional que se mantiene en el objeto

O de la forma $S \rightarrow P$, donde S denota un grupo de atributos y P denota un atributo entonces, S es un implicante en X en el objeto O.

$\text{implicante}(11,0) \Leftrightarrow \exists \text{At}(\text{dep-fun}((11,\text{At}),0))$.

Cumpliendo esta forma normal, en el conjunto relación APROBO se debe representar a través de las restricciones de cardinalidad, las dependencias funcionales EST, MAT $\rightarrow$ PROF. Eso se logra modificando la cardinalidad de PROF en APROBO es decir, $\text{cardinalidad}(\text{APROBO}, \text{PROF}, (0,1))$.

Es importante destacar la siguiente propiedad de los esquemas E-R normalizados:

PROP. Sea E1 un esquema E-R normalizado, sea E2 un esquema Relacional tal que $E2 = \text{TRADUCCION}(E1)$, donde TRADUCCION es el algoritmo de traducción de Markowitz[15], entonces E2 cumple como mínimo la 3FN.

5. TECNICAS DE NORMALIZACION DE ESQUEMAS E-R.

En el trabajo realizado, además de definir formas normales para el modelo E-R, se han establecido criterios de equivalencia y se han definido algoritmos de traducción de esquemas no normalizados a esquemas normalizados.

El mejor de estos algoritmos debido a que preserva toda la información semántica que se representaba originalmente, es el algoritmo que se basa en transformar la fórmula lógica que representa el esquema E-R en una fórmula equivalente que representa un esquema E-R normalizado, haciendo sustituciones sucesivas. Los pasos de este algoritmo son:

- 1) Traducir el esquema E-R a la conjunción de las fórmulas

lógicas del conjunto de axiomas y restricciones de integridad.

2) Sustituir cada subfórmula correspondiente a un conjunto de objetos con dependencias funcionales redundantes, por una subfórmula equivalente que corresponde al mismo conjunto de objetos pero sin las dependencias funcionales redundantes.

3) Sustituir cada subfórmula correspondiente a un conjunto entidad no normalizado, por una fórmula equivalente, que corresponde a un esquema E-R normalizado.

4) Sustituir cada subfórmula correspondiente a un conjunto relación no normalizado, por una fórmula equivalente, que corresponde a un esquema E-R normalizado.

5) Traducir la fórmula normalizada, al esquema E-R correspondiente.

Con este algoritmo se logra hacer un proceso de normalización a nivel del modelo E-R, que preserva la información que originalmente se quería representar.

6. TECNICAS DE TRADUCCION DE E-R A RELACIONAL.

Como se dijo anteriormente, no existen manejadores de bases de datos comerciales cuya herramienta de especificación es el modelo E-R. Luego es de interés la existencia de algoritmos de traducción a los modelos que si ofrecen los manejadores que tengan la propiedad de preservar en el esquema resultado, toda la semántica que originalmente se había representado.

Existen varios algoritmos de traducción de esquemas E-R a Relacional, por ejemplo el de Ullman, y el Markowitz[15,20], entre otros. En ambos existe pérdida semántica al hacer la traducción a relacional, ya que por ejemplo con el de Markowitz no se representa el límite inferior de la cardinalidad ni

tampoco, el límite superior cuando es diferente de infinito. Por otro lado, utilizando el algoritmo de Ullman, además de perder esta información, no se representa la información que depende del rol que juegan los objetos en una relación.

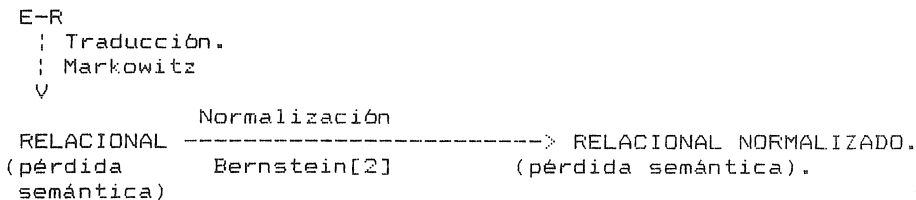
Para subsanar estas limitaciones, en el trabajo se propone un algoritmo que traduce todas las abstracciones que no se pueden representar a través de una estructura del relacional, como una restricción de integridad. Por ejemplo: la restricción que un estudiante no puede inscribir menos de 2 materias ni más de 5, que en el E-R se puede representar a través de cardinalidad(INSCRIBE,MAT,(2,5)), representarla en relacional de la siguiente manera:

$$\neg \exists \text{ est } (\{ (\text{est}, \text{mat}) \mid \text{tupla}((\text{est}, \text{mat}), \text{INSCRIBE}) \} \mid < 2 \vee 5 < \{ (\text{est}, \text{mat}) \mid \text{tupla}((\text{est}, \text{mat}), \text{INSCRIBE}) \})$$

Donde INSCRIBE es una relación del esquema relacional y $\text{tupla}(x,R)$ si "x es una instancia de la relación R".

CONCLUSIONES.

Tradicionalmente para poder definir una base de datos normalizada se tenían que seguir los siguientes pasos:

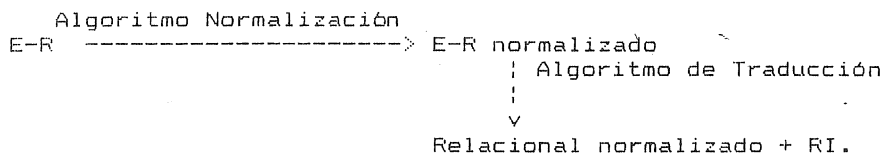


Aparte de la pérdida de información que se presentaba en cada nivel de traducción y que el diseñador conceptual no siempre podía verificar si el esquema normalizado realmente correspondía con la realidad que quería representar, se presentaban problemas

cuando no se utilizaba el modelo Relacional para implementar la BD.

Ofreciendo un Teoría de Normalización para un modelo de Datos más semántico y definiendo algoritmos apropiados de traducción, se pueden subsanar las limitaciones anteriormente presentadas.

De esta forma, para el caso en que sí se implementa la B.D con el modelo Relacional y se realiza la modelación conceptual con E-R, se puede obtener un esquema normalizado que realmente modela toda la información que originalmente se había representado. Así, se podría ver el proceso de diseño de una Base de Datos de la siguiente forma:



Además estos algoritmos pueden ser utilizados para herramientas CASE que asistan en el diseño de BD. De hecho los mismos serán utilizados en el modulo de diseño de BD de la herramienta CASE que se está desarrollando en la Universidad Simón Bolívar.

Una posible extensión a este trabajo es extender la teoría al modelo de datos Entidad Relación Extendido[18].

BIBLIOGRAFIA.

0. BAGCHI, T. y CHAUDHRI V. "Interactive Relational Database Design". Springer-Verlag. 1987.
1. BARRY, E. J. "Applied Database Logic", Vol I Fundamental Database Issues. Prentince-Hall, 1985.
2. BERNSTEIN, P. A. "Synthesizing third normal form relations from

functional dependencies". ACM Trans. Database Syst. 1, 4 (Dec.1976).

3. BERRI, C. y BERSTEIN, P. A. "Computacional problems related to the design of normal form relational schemas". ACM Trans. Database Syst. 4, 1, (Mar.1979).

4. CERI, S. y GOTTLÖB, G. "Normalization of Relations and Prolog", Communications of the ACM. 29, 6, (Jun,1986).

5. CODD, E. F. "A Relational Model of Data for Large Shared Data Banks". Communications of the ACM. 13, 6, (Jun,1970).

6. CODD, E. F. "Extending the Database Relational Model to Capture More Meaning", ACM Trans. Database Syst. 4, 4, (Dec 1979).

7. CHEN, P.P. "The Entity Relationship Model-Toward a Unified View of Data", ACM Trans. Database Syst. 1, 1, (Mar 1976).

8. DATE, C. J. "An Introduction to Database Systems", Fourth Edition 1986. Addison-Wesley.

9. DIEDERICH, J. y MILTON, J. "New Methods and Fast Algorithms for Database Normalization". ACM Trans. Database Syst. 13, 3, (Sep 1988).

10. FAGIN, R. " Multivalued dependencies and new form for relational database", ACM Trans. Database Syst. 2, 3 (Sep 1977).

11. FAGIN, R., MENDELZON, A., ULLMAN J. "A Simplified Universal Relation Assumption and its Properties", ACM Trans. Database Syst. 7,3 (Sep 1982).

12. HULL R. y KING R., "Semantic Database Modeling: Survey, Applications, and Research Issues", ACM Comp.Surv. 19, 3 (Sep 1987).

13. MAEIER, D. "Minimun Covers in the Relational Database Model" J. ACM, 27, 4 (Oct 1980).

14. MAEIER, D., SAGIV Y., YANNAKAKIS M. "On the Complexity of Testing Implications of Functional and Join Dependencies" J. ACM, 28, 4 (Oct 1981).

15. MARKOWITZ V. y SHOSHANI A. "On the correctness of representing Extended Entity Relationship structures in the Relational Model", ACM Procc. 1989.

16. PECKHAM J. y MARYANSKI F. "Semantic Data Models" ACM Comp. Surv. 20, 3 (Sep 1988).

17. SILVEIRA P. "Procedural Data Manipulation Operations for E-R Model" Conferencia Latinoamericana de Informática 1990.

18. TEORY T., YANG D., FRY J. "A Logical Design Methodology for Relational Databases Using the Extended Entity-Relationship Model". ACM Comp. Surv. 18, 2 (Jun 1986).

19. TSICHRITZIS, D. y LOCHOVSKY, F. "Data Models". 1982.

20. ULLMAN J. "Principles of Database and Knowledge-Base Systems", Vol I. 1988. Computer Science Press.