

## **Diseño de BD Relacionales. Otro enfoque.**

**René C Martínez Baños, PhD**

Empresa Telefónica de Cuba SA, Ciudad Habana, Cuba, [rmartinez@tel.etcসা.сu](mailto:rmartinez@tel.etcসা.сu)

### **Abstract**

Este trabajo es el resultado de 10 años de docencia de pregrado en la carrera Ingeniería Informática en el Instituto Superior Politécnico José A Echevarría y de otros 5 años en cursos de maestría en la propia institución. Ha sido de mucha utilidad para mejorar los niveles de conocimientos de los estudiantes en ambos niveles de enseñanza y el objetivo fundamental del mismo es mostrar cuales han sido las modificaciones realizadas y los cambios en el enfoque utilizado para lograr estos resultados.

### **Keywords**

Bases de Datos, Nuevo enfoque, Enseñanza, Relacional

### **Introducción**

En este trabajo se presenta el tema *Diseño de Bases de Datos*, con una óptica distinta a como lo hacen otros textos. El propósito que aquí se persigue es exponer un conjunto de ideas puestas en práctica tanto en la enseñanza como en el diseño profesional de Bases de Datos Relacionales.

Se incluyen términos nuevos, se modifican en cierta medida la percepción de otros y se hacen recomendaciones que rompen con determinadas tradiciones. Los términos manejados son:

1. Modelo o Enfoque Relacional.
2. Base de Datos.
3. Banco de Datos.
4. Relación o Tabla.
5. Lentes. Utilidad pedagógica.
6. Base de Datos Minimal.
7. Base de Datos Definitiva o Mejorada.
8. Interpretación de las dependencias funcionales.
9. Espacio Existencial.
10. Lenguajes para las etapas.
11. Nuevo lenguaje para describir al modelo conceptual.

Se aclara, para tranquilidad del lector, que en este trabajo se desestiman varios de los términos más usados en el tema diseño de Bases de datos relacionales usándolos con diferentes contenidos e introduciendo nuevos.

Se deja al lector concluir cuales son nuevos y cuales no.

## 1. **Modelo o Enfoque Relacional.**

En todos los textos de BD se usa el término Modelo Relacional para identificar al conjunto de conceptos, herramientas y métodos que se utilizan para el diseño de Bases de Datos Relacionales. Se discute sólo el Modelo Relacional porque los demás tienen únicamente interés histórico razón por la cual siempre se obvia mencionarlos. Excepto el Enfoque Orientado a Objetos

La primera controversia comienza aquí. Se plantea en este trabajo que lo que se dice para el modelo relacional debe decirse para lo que denominamos Enfoque Relacional pues Modelo Relacional se entiende que es un sustituto funcional construido por el hombre para determinado Sistema Objeto y que no puede existir desligado de él. Sin embargo el enfoque relacional con todas sus características si existe separado de cualquier realidad.

Conclusiones: El enfoque relacional es lo que se decía para el modelo y Modelo Relacional de algún Universo del Discurso es lo que se denomina aquí su Base de Datos Relacional.

## 2. **Base de Datos.**

El concepto **Base de Datos (BD)** está asociado indudablemente con lo que se denominará **Universo del Discurso (UD)**. La vinculación establecida entre los conceptos **BD** y **UD** está dada por el hecho de que una **BD** se diseña para un **UD** y no tiene sentido hablar de una, sin el otro.

Según Date [1] en la referencia 1.1, un Sistema de Bases de Datos es "*un Sistema computarizado de mantenimiento de registros*" y la Bases de Datos en sí misma, la caracteriza como "*un almacén electrónico rellenable*" o "*almacén para una colección de ficheros de datos computarizados*". Continúa diciendo que "*al usuario del sistema se le dan ciertas facilidades para realizar determinadas operaciones*" que enumera enseguida.

En la referencia 1.2 hace, entre otras cosas, una distinción interesante entre **dato** e **información** diciendo que **dato** es el valor almacenado e **información** su significado. Es interesante añadir que **información** también puede ser *datos procesados*.

Después de presentar otros conceptos, llega a la referencia 1.3 y se pregunta: ¿Qué es una Bases de Datos? a lo que responde: "*una colección de datos persistentes utilizada por una aplicación en una*

*organización empresarial o de servicios*". Claramente, comienza explicando la diferencia entre dato transiente y persistente comentando que los persistentes son aquellos que se almacenan en la Base de Datos y tienen un carácter menos efímeros que los datos de entrada, salida y resultados intermedios.

Otros autores hacen una descripción más o menos semejante cuando tratan el término BD.

Lamentablemente no se coincide con ellos y se plantea que ninguna de esas ideas deja claro en la mente de quien desea construir una BD ¿qué es lo que tiene que hacer?, tampoco a qué debe llegar ni cómo hacerlo.

Parece y de hecho se ha probado un número de veces no despreciable que una descripción como la que sigue es más efectiva.

El concepto de BD se plantea como un *Modelo* de cierto *UD*. Está claro que por razones de espacio no se puede transitar por todo el camino que lleva a esta conclusión. Baste decir que Modelo en el mundo de los datos (como sustituto funcional de determinada realidad o Sistema Objeto de Estudio), es el contenido esencial del concepto de BD. Lo acompañan también propiedades como:

- según el interés del observador
- sin redundancias dañinas
- con carácter permanente
- actualizable y
- utilizando un enfoque Relacional.

Se excluye por el momento el enfoque Orientado a Objetos por declararnos incapaces de formalizarlo con todo rigor. De esta manera se dice que una Base de datos es:

Un modelo (*Reflejo fiel*) en el mundo de los datos de cierto UD según el interés del observador, sin redundancias dañinas, con carácter permanente, actualizable y construido utilizando un enfoque Relacional.

### 3. Banco de Datos.

Conjunto de relaciones que modelan a los estados pasados (estáticos) de ciertas entidades de interés que forman parte del UD. Habitualmente están en 1ªFN por cuanto **no se pueden actualizar**.

Un modelo completo del UD incluye a la Base de datos para modelar su parte dinámica (Estado actual) y un Banco de datos para modelar los estados pasados de interés (Historia del UD)

La enseñanza de diseño de bases de datos no puede obviar la discusión de estos conceptos.

#### 4. Relación y Tabla.

Una Relación se define formalmente como:

$\mathbf{R} \subseteq \mathbf{D}_1 \times \mathbf{D}_2 \times \dots \times \mathbf{D}_n$  donde:

.- $n$  es el grado de la relación o tabla y la cantidad de **dominios** que participan en el producto cartesiano, definen un **dominio compuesto** denominado de ahora en adelante el *espacio existencial de la relación*. (los **dominios** no son soportados por la mayoría de los SGBD que se dicen relacionales)

.- $\mathbf{D}_i$  es el  $i$ ésimo **dominio** (no necesariamente distintos) del espacio de **R**.

Ese producto define al dominio compuesto para el conjunto de tuplas **t** de la forma:

$\mathbf{t} = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n)$  de modo que  $(\mathbf{d}_1 \in \mathbf{D}_1, \mathbf{d}_2 \in \mathbf{D}_2, \dots, \mathbf{d}_n \in \mathbf{D}_n)$ .

Sean:

.-  $\mathbf{t}_i$  una tupla de ese espacio *o universo*.

.-**M** la extensión del producto cartesiano de los dominios.

.-**E** = { $\mathbf{t}_i / i = [1, M]$  }

.-**P** un predicado que caracteriza los miembros de **E** que pueden formar parte de **R**.

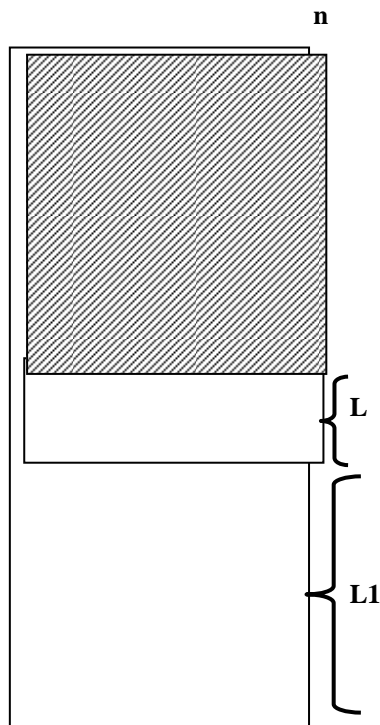
Se dice entonces que:

$$\mathbf{R} = \{ \mathbf{t} / \mathbf{P}(\mathbf{t}) \wedge (\mathbf{t} \in \mathbf{E}) \}$$

Se interpreta diciendo que **R** es una *variable* definida en cierto espacio existencial **E** y su *valor* es el conjunto de tuplas **t** que satisfacen el predicado **P** en cierto momento del tiempo.

Resulta imperativo incluir algunas consideraciones en este punto, que van a distinguir la presentación del elemento teórico básico del modelo relacional (Relación) que se hace en este libro, del que se encuentra en los textos clásicos.

Si se usa un SGBD que no soporta dominios, las *relaciones* con las que trabaja residen en un espacio compuesto por el **producto cartesiano de n tipos de datos**. Este espacio es mucho mayor que el definido por el **producto cartesiano de los dominios** asociados (ver figura).



El rectángulo mayor representa al espacio definido por el producto de n tipos de datos y el menor por el producto de n dominios. Las tuplas válidas serán las que satisfagan la definición  $R(A, F, \mathbf{L})$  es decir, área sombreada.

- $\mathbf{L}$  serán los predicados que invalidan a las tuplas ilegales del espacio existencial de R.
- $\mathbf{L1}$  son los predicados que tendrán que incluirse, como restricciones adicionales, para que el SGBD opere con tablas (no con relaciones) capaces de comportarse como si lo fueran.  $\mathbf{L} + \mathbf{L1}$  serán las restricciones de integridad necesarias para que el sistema que se diseña en un SGBD sin dominios.
- F serán todas las dependencias funcionales que se sostienen en el UD

## 5. Lentes. Utilidad pedagógica.

La BD, como **imagen** en el mundo de los datos de un sistema, se '*ve*' a través de tres lentes. Hasta el momento sólo se ha mencionado al que refleja el interés del usuario.

Se puede añadir un **segundo** lente encargado de reflejar **la calidad del diseñador**. Hay que considerar que cada diseñador puede lograr una **imagen** distinta del **UD**, en razón de su más o menos acabada formación académica, mayor o menor experiencia, etc. Se hace notar que este lente es el de más

fácil eliminación por cuanto las razones de su existencia dependen enteramente del hombre y pueden ser equilibradas.

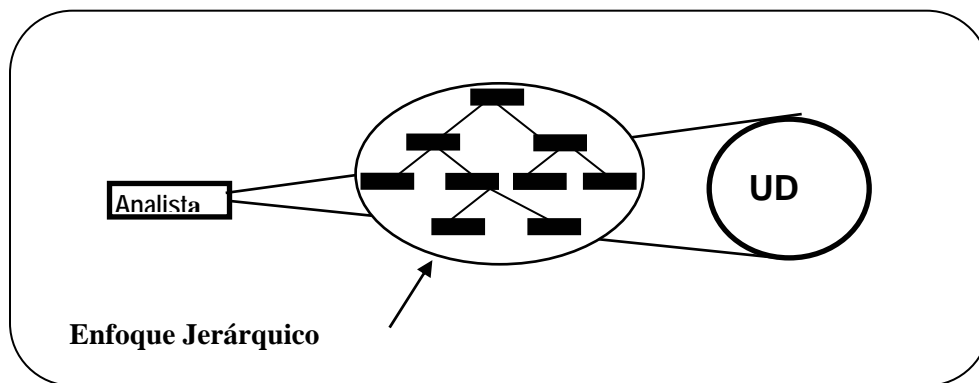
Existe un **tercer** lente entre el **UD** y los **ojos** del diseñador, asociado al Enfoque de Diseño que se decide emplear para construir la BD. Considerando los tres enfoques clásicos, la **imagen** del **UD** en el mundo de los datos puede tomar la forma de:

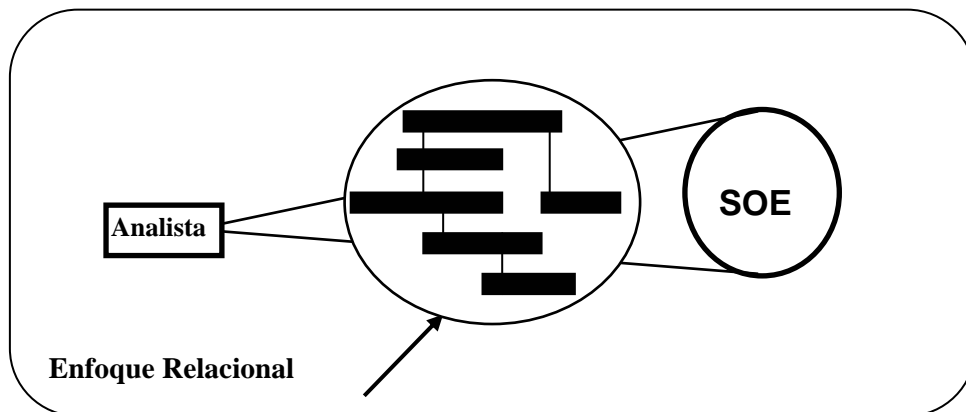
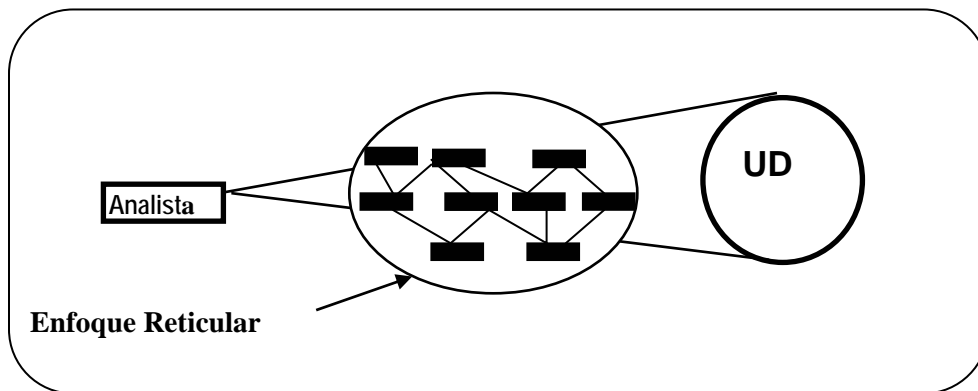
- un árbol en el caso que se emplee el Jerárquico.
- una red en el caso que se emplee el Reticular y
- un conjunto de tablas enlazadas en el caso que se emplee el Relacional.

De esta manera, si no se toman medidas para eliminar o disminuir sus efectos, la **BD** como imagen de un **único UD** en el mundo de los datos será diferente.

De los tres lentes mencionados:

- el que tiene que ver con la **calidad del analista**, se elimina nivelando conocimientos, experiencia y usando todos la misma metodología.
- el que tiene que ver con **el enfoque de diseño**, se puede eliminar si se adopta un sólo enfoque. En la actualidad, ese es el caso del Relacional.
- el que tiene que ver con **el interés del usuario**, no se puede eliminar pues con él se decide **conceptualmente** qué informaciones se deben registrar, cuáles son los vínculos o nexos que existen entre ellas y con cuáles propósitos se necesitan.





## 6. Base de Datos Minimal.

Cuando se diseña una BD a partir de la cubierta minimal de las Dependencias Funcionales y Multievaluadas que se sostienen en el espacio existencial del UD, se logra lo que se denomina Base de datos minimal, es decir aquella formada por un conjunto de relaciones en 5FN que representan con total fidelidad al UD del cual es un reflejo.

## 7. Base de Datos Definitiva o mejorada.

La Base de Datos Definitiva o Mejorada se le dice a aquel conjunto de Relaciones que se eligen para formar la BD y es el que se declara en el SGBD elegido, estén o no en 5FN. En muchas ocasiones la BD minimal no es todo lo efectiva que se quisiera para algunas consultas muy frecuentes en lo que a consumo de tiempo se refiere. En esos casos se ‘mejora’ la BD desnormalizandola. El proceso de desnormalización también requiere de la máxima atención y no es posible tratarlo en este trabajo pues al respecto hay escritas 100 páginas más.

## 8. Dependencias funcionales y Espacio Existencial.

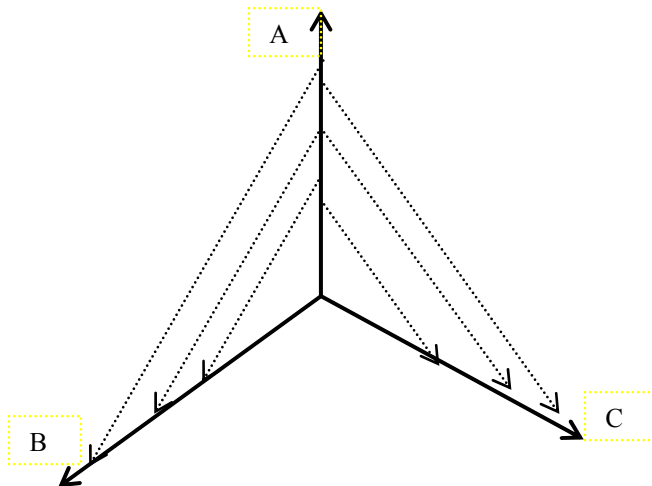
El estudio de las dependencias funcionales, polivalentes y multievaluadas que se sostienen en el UD tiene la mayor importancia y en muchas ocasiones este se soslaya. De los textos revisados sólo en el de Ullman se les da la atención requerida. El término dependencias funcionales polivalentes debe llamar la atención pues no se ve en los textos clásicos.

Se dice que una dependencia es funcional cuando un elemento de X tiene uno y solo uno asociado de Y. Se dice que es polivalente cuando un elemento de X tiene más de uno pero en todos los casos la misma cantidad de valores en Y para cualquier X. Esto es útil para evitar relaciones muchos a muchos cuando no son necesarias.

El estudio de las dependencias funcionales es esencial si se quiere entender como es preciso qué se está haciendo cuando se diseña una BD relacional. Aquí se expone únicamente el papel que juegan al caracterizar el espacio existencial.

Es frecuente ver definida una relación con la expresión  $R(A)$  y nada más lejos de la realidad. Tiene que ser definida al menos usando  $R(A, F)$  aun cuando lo correcto es  $R(A, F, L)$  dejando claro no solo a los dominios que participan (A) y las dependencias que se sostienen (F) sino además los predicados (L) que limitan a los valores legales de R.

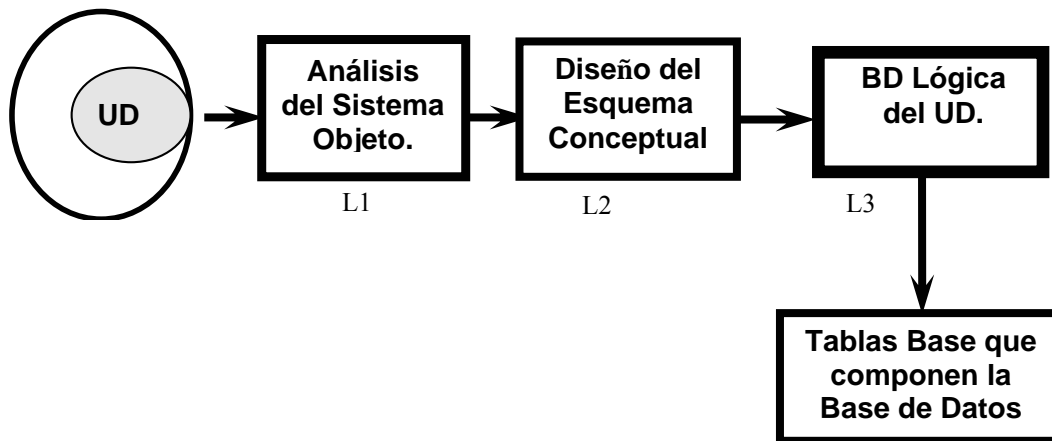
Una forma simple de mostrar el papel de F se puede apreciar en los siguientes ejemplos:



Este es un espacio definido en 3 dominios (A, B, C) y con  $F = \{A \rightarrow B, A \rightarrow C\}$  Sin embargo, otro espacio definido también en los mismos 3 dominios (A, B, C) con un valor para  $F = \{AB \rightarrow C\}$  es totalmente diferente. En el primero, basta conocer un valor de A para conocer a los valores correspondientes de B y de C y en el segundo es necesario conocer a un par AB para conocer el valor de C correspondiente. Si se emplea  $R(A)$  para caracterizar a R en lugar de  $R(A, F, L)$  se comete un error grave. Omitir a F es imperdonable.



## 9. Lenguajes para las etapas.



Desde el punto de vista docente, ha reportados grandes resultados plantear que el modelo del UD (BD) que se logra al final del proceso esquematizado arriba es producto de sucesivas traducciones del mismo modelo expresado en diferentes lenguajes.

En primer término se expresa al modelo en lenguaje natural (L1) cuando se concluyen las especificaciones luego del análisis del sistema. Que este paso se da en varias ‘vueltas’ de un proceso iterativo, se admite.

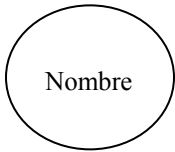
En segundo término se expresa el mismo modelo en un lenguaje gráfico obtenido mediante una ‘traducción’ del modelo inicial (L2). Este lenguaje gráfico es por excelencia conocido como DER que permite hacer lo que se llama Modelo Conceptual del UD.

Y en última instancia se expresa el modelo en su forma final (L3) como un conjunto de Relaciones (Tablas Base) que forman la BD del UD. Este último lenguaje es común para cualquier variante de lenguaje gráfico, de la misma forma que lo es el binario para cualquier compilador.

## 10. Nuevo lenguaje para describir al modelo conceptual.

Este lenguaje gráfico no es un lenguaje formal, como lenguaje se describe enumerando los símbolos admisibles y la semántica de cada uno, aislados y en conjunción con otros. Tiene características semejantes al D.E.R. y se usa para los mismos propósitos. A saber: Construir un esquema conceptual de la BD.

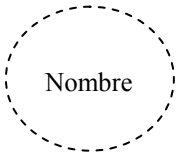
Símbolos admisibles.



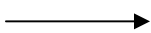
Se utiliza un círculo para representar a las Entidades Naturales o Regulares.



Se utiliza para representar a las propiedades de las entidades.



Se utiliza un círculo con línea punteada para representar a las Entidades Débiles



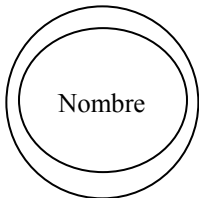
Representa a un enlace simplemente evaluado entre una Entidad Regular o Débil (o una entidad Construida), con sus atributos u otra Entidad. Puede estar etiquetado.



Representa a un enlace multi-evaluado entre una Entidad Regular o Débil o Construida, con otra Entidad de cualquier tipo.

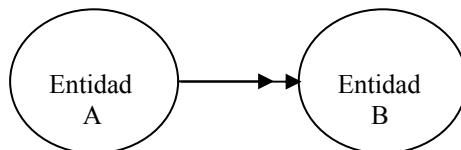


Representa a una bandera de dominio indicadora de comunidad de dominios entre grupos de atributos.

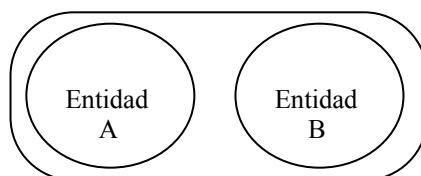


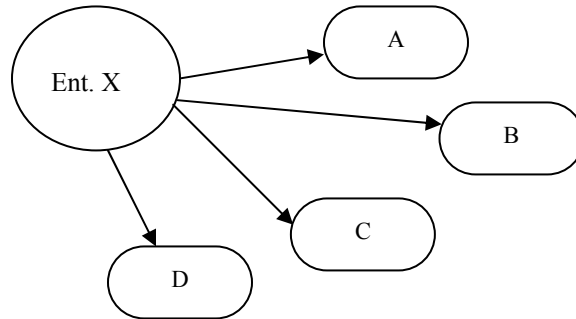
El círculo exterior (dos o más) se usan para aislar a un nodo cuando forma parte de varias cadenas, o cuando tiene que ser distinguido como especialización.

Sinónimos de los Símbolos admisibles.



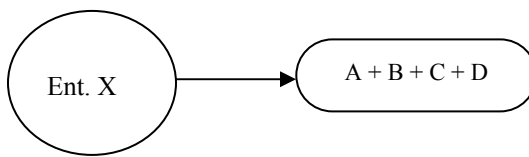
Este esquema anterior es equivalente al siguiente.





Este esquema está indicando que la entidad X tiene 4 propiedades A, B, C y D

El siguiente esquema indica lo mismo de forma más compacta.



### Conclusiones.

En este trabajo se han expuesto, utilizando el mínimo de elementos, ideas nuevas, términos nuevos y modificados al menos en cuanto a su interpretación. El origen de todo se encuentra en la necesidad de mejorar la preparación que reciben los estudiantes de informática para enfrentar el diseño de Bases de Datos relacionales. Se logra llevarlos a un nivel de confianza superior cuando se enfrentan a un ejercicio académico y ¿por qué no? Profesional. Adquieren mayor soltura manejando términos del área y se comunican más fácilmente con especialistas de la materia.

Lamentablemente no he tenido espacio para ser más explícito con todos estos elementos, sobre todo con el último, el Lenguaje gráfico para representar al modelo conceptual. Me atrevo a decir que tiene mejor poder semántico que el DER pues en el DER no es posible, sin modificaciones, representar especializaciones de agregaciones y en este si.

La experiencia que he tenido en las clases y el trabajo profesional con este lenguaje ha sido muy satisfactoria y alentadora.

Tengo la conciencia clara respecto al hecho de que el DER cuenta con varios CASE para generar tablas a partir del esquema que UD puede ‘dibujar’ en el ambiente de tales herramientas y este no tiene ninguna. Entre ellos el Edwin y el EREstudio. He trabajado con ambos y me encontrado con situaciones en que he tenido que hacer las tablas a mano.

Es mi deseo que este esfuerzo sea de utilidad para una persona. Si al final es útil para más de una, mucha más alegría.

Es interesante destacar que del DER se extraen tablas en 3FN para concluir las a mano y de este lenguaje se extraen tablas en 5FN.

## **Bibliografía**

A. Silberschatz, H.F. Korth y S. Sudarsham: "DataBase System Concepts", 3 Ed. NY, McGraw-Hill, 1977.

C.J.Date, "Introducción a los Sistemas de Bases de Datos". Séptima edición, Ed. Félix Varela. La Habana. 2003.

E.F. Codd. "A Relational Model of Data for Large Shared Data Banks" CACM 13 N° 6 junio 1970.

E.F. Codd. "Relational Database: A Practical Foundation for Productivity", CACM 25, N° 2 febrero 1982.

J.D.Ullman."Database and Knowledge-base Systems", Computer Sciece Press, 1988, 1989.

R. Fagin: "Functional Dependencies in a Relational Database and propositional Logic" IBM J. R&D 21, N° 6 noviembre 1977.

## **Informacion Biografica**

Dr. René C. MARTÍNEZ BAÑOS. El Dr. Martínez Baños es Ingeniero Industrial con Master en Diseño de Sistemas y PhD en Investigación de Operaciones. Ha sido profesor en el Instituto Superior Politécnico José Antonio Echeverría (ISPJAE) en Cuba desde 1966. Ha enseñado cursos de postgrado, mayormente de Informática, en UAGRM-Bolivia, UAS-Mexico, UNOPAR-Brazil, Universidad de Camagüey-Cuba, UCICA-Colombia, CENTERSYSTEMS-Colombia y UMSA-Bolivia.