

# Patrones de Reglas del Negocio para el Enriquecimiento de un MER

Viviana E. Ferragine

Laura C. Rivero

INTIA - Facultad de Ciencias Exactas – U.N.C.P.B.A. Tandil, Buenos Aires, Argentina

{vferra, lrivero} at exa.unicen.edu.ar

**Resumen:** El principal objetivo del diseño de una base de datos es capturar la semántica estructural de los sistemas de información, especificando objetos relevantes para la aplicación y sus interconexiones. Esto puede lograrse usando un modelo de datos semántico y traduciendo luego tal especificación en una definición de base de datos relacional. Este proyecto tiene como meta identificar y formalizar un conjunto completo de patrones de reglas del negocio que puedan plasmarse en un Modelo de Entidades y Relaciones (*MER*) convencional utilizando las construcciones básicas universalmente aceptadas. Para los casos en los que ésto no es posible, se sugieren formas alternativas de especificación sin perder semántica estructural, o comportamiento. Este proyecto se centra en reglas del negocio expresables mediante dependencias de comparación de valores, especialmente dependencias de inclusión y de exclusión (primitivas). Su estudio se extenderá a reglas sustentadas por dos tipos derivados: dependencias de igualdad y de superposición parcial.

## 1. Introducción

El modelo conceptual de una base de datos relacional generalmente está descrito por una de las extensiones del muy difundido Modelo de Entidades y Relaciones (*MER*) [3]. El *MER* tiene limitaciones expresivas, las que paradójicamente han contribuido a su amplia aplicabilidad, a su formalización y simplicidad de uso [2]. Actualmente, se requiere dar soporte a aplicaciones más complejas capturando más semántica del dominio, y esto sólo puede lograrse si la misma se recogió y representó en el modelo conceptual. En esta dirección, cualquier adición a un modelo de datos, el *MER* u otro, no sólo debe estar muy bien motivada, sino que debe ser minimal y debe producir un importante impacto en el poder expresivo. En el caso del *MER* solo pueden combinarse atributos, entidades y relaciones de ciertas maneras, no libremente; existe una restricción inherente que sólo admite relaciones entre entidades, no entre entidades y relaciones o bien relaciones entre sí [3], [4]. Existen numerosas propuestas concibiendo importantes mejoras en el nivel semántico, y han planteando la forma de representarlas en el Diagrama de Entidades y Relaciones (*DER*). En [15] se considera la generalización de supertipos y subtipos; la relación de rol total reflejando la participación obligatoria de una entidad en relación con otra; la restricción de exclusión para jerarquías exclusivas y la restricción de subconjunto o dependencia de inclusión. La propuesta de [8] da lugar al *MER Extendido*, incorporando los conceptos de clase/subclase y herencia de tipos, especialización y generalización conjuntamente con sus restricciones y categorías. Existen otros modelos con distinto punto de vista u orientación, como Merise [14], ORM [9], o UML [16] en los que es posible modelar además de las restricciones descriptas, otras que afectan a los tipos de relaciones y a sus ejemplares.

Las dependencias de datos son restricciones en las que valores en tuplas de relaciones en una base de datos dependen de igualdades o desigualdades de otros valores [10]. Estas han sido tratadas, formalizadas y estudiadas, con diferente grado de profundidad y desde diferentes enfoques. Uno de ellos es el que se realiza desde el punto de vista de las reglas del negocio. La captura de estas reglas mediante el modelado de los datos (como parte de la definición de las entidades, las relaciones, sus cardinalidades y los atributos) aún no es una tarea sencilla ya que generalmente estas reglas son complejas, tienen una estructura arbitraria, cambian frecuentemente y son difíciles de forzar. Es innegable la necesidad de investigar en aspectos tales como la captura y especificación de estas restricciones desde su expresión en el UdeD hasta plasmarla en una base de datos [1], aunque desde los inicios de esta actividad los avances en este sentido han sido más que significativos. Baste comparar el poder expresivo del *MER* original de Chen [3] con los actuales UML [16], ORM [9], Merise [14], o con diferentes extensiones del *MER* [15], [8], [6], y otros tal vez menos difundidos.

En el modelo relacional las dependencias de datos primitivas son las dependencias funcionales y las dependencias de inclusión. La caracterización de las dependencias de inclusión se puede extender a dependencias de exclusión [15], [7], [1] y en general a una clase que las engloba con otras de tipo similar: dependencias de comparación de conjuntos. Numerosas reglas del negocio pueden expresarse mediante estos formalismos.

Una gran variedad de restricciones pueden plasmarse naturalmente en el *MER Extendido*: restricciones de claves, de cardinalidad y de integridad referencial, pero otras como algunas dependencias de inclusión y las dependencias de exclusión, no pueden ser modeladas directamente. Existen varias notaciones para representar los conceptos del *MER* y sus extensiones, pero no hay estándares. Asimismo, el poder expresivo de los *DER* o *DER Extendido*, es más amplio que el de las diversas metodologías CASE, ya que la mayoría de ellas, sobre todo las comerciales, restringen el modelado a relaciones binarias, o limitan las opcionalidades permitidas, u obligan a reificar relaciones para plasmar aspectos relevantes. Por otra parte, los sistemas de la vida real son mucho más complejos que lo que uno puede representar en el *MER* o en alguna de sus extensiones, por lo tanto para diseñar una base de datos es crucial considerar exhaustivamente la documentación procedente de la fase de especificación de requisitos. Ésta información, de otra forma podría perderse en el diseño, aunque a veces es difícil de interpretar porque está escrita generalmente en lenguaje natural y en forma ambigua o imprecisa ya que no existen métodos (semi)formales establecidos para incorporarla en ésta etapa.

El objeto de este proyecto es investigar las dependencias de comparación de conjuntos de valores, haciendo especial hincapié en las *dependencias de inclusión* y de *exclusión* en sus aspectos tanto formales como semánticos. En este contexto se realiza un catálogo de diferentes casos, proponiendo formas alternativas de representación y una justificación de su origen como soporte a reglas del negocio en el dominio de la aplicación. Dos dependencias derivadas de aquéllas -de *igualdad* y de *superposición parcial*, serán estudiadas aprovechando el marco teórico que brinda el análisis de las dos primeras. Este proyecto contribuye con la representación de una serie de patrones, que sí pueden plasmarse en un *MER* convencional. Para aquellos casos en los que no es posible, se sugieren especificaciones alternativas.

Esta presentación está organizada de la siguiente manera: la sección 2 presenta la formalización de los conceptos involucrados; la sección 3 incluye una tabulación de casos para el análisis de las dependencias de inclusión y exclusión, ilustrando el problema con ejemplos de motivación; finalmente la sección 4 contiene conclusiones, resultados esperados y trabajos futuros.

## 2. Formalización

Un esquema relacional queda definido por  $R = \langle R, D \rangle$ , donde  $R = \{R_1, R_2, \dots, R_n\}$  representa el conjunto de esquemas de relaciones y  $D$  el conjunto de restricciones.  $D$  incluye las *dependencias funcionales*, de *inclusión* [5], [6], de *exclusión* [1], de *igualdad* y de *superposición parcial*.

### 2.1. Dependencias de comparación de valores

La existencia de (conjuntos de) atributos compatibles, justifica efectuar comparaciones entre ellos. Las restricciones de comparación de conjuntos agrupan dos tipos básicos de dependencias: de *inclusión* y de *exclusión*. Dos tipos adicionales, derivados de éstos son las dependencias de *igualdad* y de *superposición parcial*. Esta última no ha sido tratada como tal en la literatura especializada, aunque se ha analizado el problema de la superposición parcial en bajo el nombre de correlación de datos en el marco de un proceso de reingeniería de bases de datos.

#### 2.1.1. Dependencias de Inclusión (*di*)

Una dependencia de inclusión (*di*) se define como la existencia de (conjuntos de) atributos en una relación, cuyos valores deben ser un subconjunto de los valores de (conjuntos de) atributos compatibles, respectivamente en otra relación [7], [8]. Otros autores indican la posibilidad de que ambas relaciones pueden ser la misma [5].

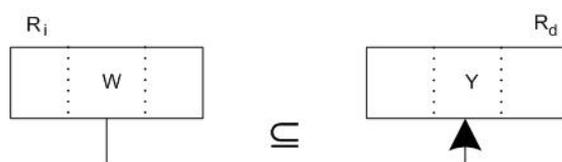


Fig. 1 : Notación gráfica para dependencias de inclusión

Formalmente una *di* es una expresión  $R_i.W \overset{I}{\mid} R_d.Y$  donde  $R_i$  y  $R_d$  son nombres de relaciones (posiblemente el mismo);  $W$  e  $Y$  son conjuntos de atributos compatibles. Cuando  $Y$  es la clave primaria de  $R_d$  la *di* es basada-en-clave y se denomina *restricción de integridad referencial (rir)* o simplemente *referencia*. En este caso  $W$  constituye una *clave extranjera* para  $R_i$ .

### 2.1.2. Dependencias de Exclusión (*de*)

Análogamente con las *di*, una dependencia de exclusión (*de*) se define como la existencia de (conjuntos de) atributos en una relación, cuyos valores no pueden pertenecer al subconjunto de los valores de (conjuntos de) atributos compatibles en otra relación (o la misma), respectivamente [1].

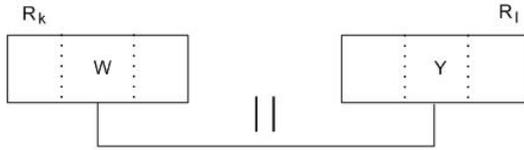


Fig. 2 : Notación gráfica para dependencias de exclusión

Una *de* es una expresión  $R_k.W \parallel R_l.Y$  donde  $R_k$  y  $R_l$  son nombres de relaciones (posiblemente el mismo);  $W$  y  $Y$  son conjuntos de atributos compatibles. Aquí no existen términos izquierdo y derecho de la dependencia, dado que la exclusión se da en ambos sentidos.

### 2.1.3. Dependencias de Igualdad (*dig*) y de Superposición Parcial (*dsp*)

Una *dependencia de igualdad (dig)* resulta de establecer sendas dependencias de inclusión inversas. Esto implica la existencia de (conjuntos de) atributos en una relación, cuyos valores deben ser los mismos que los correspondientes a (conjuntos de) atributos compatibles, en otra relación (o la misma). Razones de simetría indican una última dependencia de comparación, que denominaremos *dependencia de superposición parcial (dsp)*. Se satisface cuando el conjunto de instancias de atributos en una relación  $R_i$  (sean  $R_i.W$ ) y los correspondientes compatibles de otra relación  $R_j$  (sean  $R_j.Y$ ) cumplen con  $R_i.W \cap R_j.Y \neq \emptyset \wedge R_i.W - R_j.Y \neq \emptyset \wedge R_j.Y - R_i.W \neq \emptyset$ .

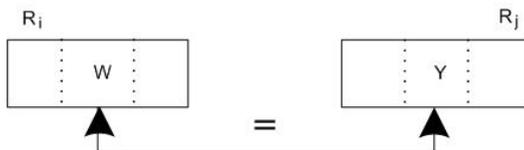


Fig. 3 : Notación gráfica para dependencias de igualdad

Una *dig* denotada como  $R_i.W = R_j.Y$  se satisface en un esquema relacional, si se satisfacen que:  
 $R_i.W \subseteq R_j.Y \wedge R_j.Y \subseteq R_i.W$   
 donde  $W$  e  $Y$  son conjuntos de atributos compatibles

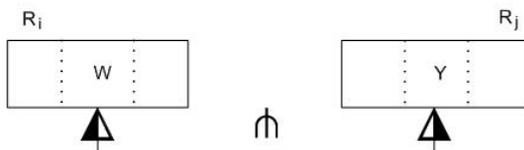


Fig. 4 : Notación gráfica para Dependencias de Superposición Parcial

Una *dsp* se denotará  $R_i.W \cap R_j.Y$  y tiene un significado equivalente a:  
 $R_i.W \not\subseteq R_j.Y \wedge R_j.Y \not\subseteq R_i.W \wedge R_i.W \cap R_j.Y \neq \emptyset$   
 donde  $W$  e  $Y$  son conjuntos de atributos compatibles

## 3. Análisis de las *dis* y *des*

En los últimos años la investigación de las *dis* ha categorizado muchos de sus aspectos. Un marco de referencia para su estudio puede encontrarse en [13], como así también el estudio de su caso particular, las restricciones de integridad referencial (*rir*). Adicionalmente un análisis de su posible origen desde una perspectiva semántica se puede encontrar en [11].

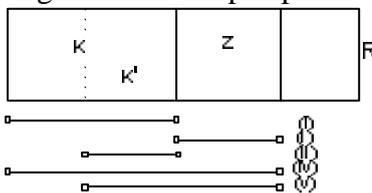


Fig. 5: correlación de claves primarias y extranjeras

Para el análisis y la comprensión semántica de las *dis* se ha seguido lo expuesto en [12] analizando los atributos involucrados, independientemente del dominio de aplicación, y tomando en cuenta las diferentes ubicaciones de los términos derecho e izquierdo de la dependencia, en relación con la disposición de la clave primaria<sup>1</sup>.

Si  $W$  es un conjunto de atributos de  $R$ ,  $K$  la clave primaria de  $R$  y  $Z$  un subconjunto de atributos secundarios, los cinco tipos o correlaciones posibles se ilustran en la Figura 5: I)  $W \equiv K$ ; II)  $W \equiv Z$ ; III)  $W \equiv K'$ , siendo  $K'$  un subconjunto estricto de  $K$ ,  $K' \neq \emptyset$ ; IV)  $W \equiv K \cup Z$ ; y finalmente V)  $W \equiv K' \cup Z$ ,  $K' \neq \emptyset$ . En todos los casos  $Z \neq \emptyset$

Teniendo en cuenta estas disposiciones, pueden obtenerse 25 casos posibles de pares  $\langle R_i[W_i], R_d[W_d] \rangle$ <sup>1</sup>, detallados en la Tabla 1. Análogamente, realizando el mismo análisis para las *des* donde sea  $W$  (conjunto de atributos de  $R$ ),  $K$  (clave primaria de  $R$ ) y  $Z$  (subconjunto de atributos secundarios), pueden obtenerse también 25 casos posibles de pares  $\langle R_k[W_k], R_l[W_l] \rangle$ . Teniendo en

<sup>1</sup> Subíndices 'i' y 'd' representan 'izquierdo' y 'derecho' respectivamente.

cuenta que la dependencia de exclusión se da en ambos sentidos, solo tiene interés realizar el análisis de la mitad de la matriz, para el resto de los casos el estudio es equivalente. La Tabla 2 muestra los casos a analizar y la equivalencia para el resto (celdas sombreadas y sin sombrear).

$W_i \backslash W_d$	I) Clave ( $K_d$ )	II) No Clave ( $Z_d$ )	III) Parte de Clave ( $K_d^1$ )	IV) Clave + No Clave ( $K_d \cup Z_d$ )	V) Parte de Clave + No Clave ( $K_d^1 \cup Z_d$ )
I) Clave ( $K_i$ )	1. $K_i \ll K_d$	6. $K_i \subseteq Z_d$	11. $K_i \subseteq K_d^1$	16. $K_i \subseteq K_d \cup Z_d$	21. $K_i \subseteq K_d^1 \cup Z_d$
II) No Clave ( $Z_i$ )	2. $Z_i \ll K_d$	7. $Z_i \subseteq Z_d$	12. $Z_i \subseteq K_d^1$	17. $Z_i \subseteq K_d \cup Z_d$	22. $Z_i \subseteq K_d^1 \cup Z_d$
III) Parte de Clave ( $K_i^1$ )	3. $K_i^1 \ll K_d$	8. $K_i^1 \subseteq Z_d$	13. $K_i^1 \subseteq K_d^1$	18. $K_i^1 \subseteq K_d \cup Z_d$	23. $K_i^1 \subseteq K_d^1 \cup Z_d$
IV) Clave + No Clave ( $K_i \cup Z_i$ )	4. $K_i \cup Z_i \ll K_d$	9. $K_i \cup Z_i \subseteq Z_d$	14. $K_i \cup Z_i \subseteq K_d^1$	19. $K_i \cup Z_i \subseteq K_d \cup Z_d$	24. $K_i \cup Z_i \subseteq K_d^1 \cup Z_d$
V) Parte de Clave + No Clave ( $K_i^1 \cup Z_i$ )	5. $K_i^1 \cup Z_i \ll K_d$	10. $K_i^1 \cup Z_i \subseteq Z_d$	15. $K_i^1 \cup Z_i \subseteq K_d^1$	20. $K_i^1 \cup Z_i \subseteq K_d \cup Z_d$	25. $K_i^1 \cup Z_i \subseteq K_d^1 \cup Z_d$

Tabla 1. Tipos de dependencias de inclusión Referencias:  $K_*$ (Clave);  $Z_*$ (atributos no-clave);  $K_*^1$ (subconjunto estricto de  $K_*$ );  $*$  = i, d

$W_k \backslash W_i$	I) Clave ( $K_i$ )	II) No Clave ( $Z_i$ )	III) Parte de Clave ( $K_i^1$ )	IV) Clave + No Clave ( $K_i \cup Z_i$ )	V) Parte de Clave + No Clave ( $K_i^1 \cup Z_i$ )
I) Clave ( $K_k$ )	1. $K_k \parallel K_i$	6. $\equiv$ Caso 2	11. $\equiv$ Caso 3	16. $\equiv$ Caso 4	21. $\equiv$ Caso 5
II) No Clave ( $Z_k$ )	2. $Z_k \parallel K_i$	7. $Z_k \parallel Z_i$	12. $\equiv$ Caso 8	17. $\equiv$ Caso 9	22. $\equiv$ Caso 10
III) Parte de Clave ( $K_k^1$ )	3. $K_k^1 \parallel K_i$	8. $K_k^1 \parallel Z_i$	13. $K_k^1 \parallel K_i^1$	18. $\equiv$ Caso 14	23. $\equiv$ Caso 15
IV) Clave + No Clave ( $K_k \cup Z_k$ )	4. $K_k \cup Z_k \parallel K_i$	9. $K_k \cup Z_k \parallel Z_i$	14. $K_k \cup Z_k \parallel K_i^1$	19. $K_k \cup Z_k \parallel K_d \cup Z_d$	24. $\equiv$ Caso 20
V) Parte de Clave + No Clave ( $K_k^1 \cup Z_k$ )	5. $K_k^1 \cup Z_k \parallel K_i$	10. $K_k^1 \cup Z_k \parallel Z_i$	15. $K_k^1 \cup Z_k \parallel K_i^1$	20. $K_k^1 \cup Z_k \parallel K_d \cup Z_d$	25. $K_k^1 \cup Z_k \parallel K_d^1 \cup Z_d$

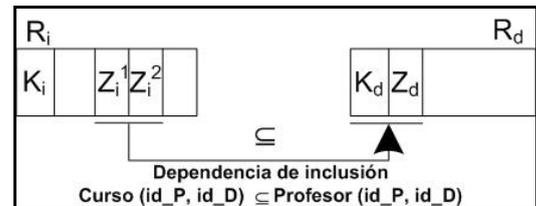
Tabla 2. Tipos de dependencias de exclusión Referencias:  $K_*$ (Clave);  $Z_*$ (atributos no-clave);  $K_*^1$ (subconjunto estricto de  $K_*$ );  $*$  = i, d

### 3.1. Ejemplos de Motivación

Analizando el esquema de relación y la regla del negocio que se presenta en el Ejemplo 2 en [2] desde el punto de vista de su semántica y tomando en cuenta las consideraciones realizadas en [12] para las *dis*, se tiene lo siguiente:

Profesor (id\_P, NombreP, id\_D)  
Curso (id\_C, Descripción, id\_D, id\_P)  
Departamento (id\_D, NombreD)  
**Esquemas de Relación**

Los profesores solo pueden dar cursos ofrecidos por los departamentos a los cuales ellos pertenecen.  
**Regla en lenguaje natural**



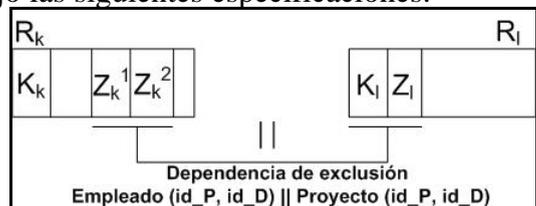
Analizando ambos términos de la dependencia se tiene que el izquierdo está compuesto por atributos no-clave (Tipo II) y el derecho por atributos clave más atributos no-clave (Tipo IV), con lo que se está frente a una dependencia de inclusión catalogada como Caso 17 en la Tabla 1.

De acuerdo a lo analizado en [12], si existe la regla y la participación es total, tanto para la relación que vincula Curso con Profesor, como para la que vincula Curso con Departamento, se tiene que ésta última puede ser convertida utilizando la Regla de Pullback [10]. Si se analizan las dependencias funcionales del ejemplo, se tiene que en la relación Profesor  $Id_P \rightarrow id_D$ , tomando en cuenta la *di* Curso (id\_P, id\_D)  $\subseteq$  Profesor (id\_P, id\_D), entonces en Curso dicha dependencia también debe mantenerse, es decir en la relación Curso también  $Id_P \rightarrow id_D$ , con lo que esta relación no se encontraría en  $\mathcal{3FN}$ . Eliminando esas transitividades la relación que existe entre Curso y Departamento puede ser eliminada, ya que puede ser deducida a través de las relaciones Curso y Profesor, Profesor y Departamento. Véase la heurística de transformación en [12].

Estudiando un ejemplo análogo al anterior para una *de*, bajo las siguientes especificaciones:

Empleado(id\_E, NombreE, id\_P, id\_D)  
Proyecto (id\_P, Denominación, id\_D)  
Departamento (id\_D, NombreD)  
**Esquemas de Relación**

Los empleados no pueden revisar proyectos del departamento para el cual trabajan.  
**Regla en lenguaje natural**



El ejemplo responde al mismo patrón del ejemplo anterior, Caso 17, pero en este caso la relación entre Proyecto y Empleado no puede ser eliminada; se debe conservar para poder verificar la exclusión. Este tipo de dependencias no ha sido axiomatizada por lo tanto su análisis se realiza caso por caso y cada patrón propuesto resulta de este análisis. Este caso no puede representarse con las herramientas de un *MER* convencional y no existe forma específica de expresarlo en SQL salvo por medio de cláusulas *check*, al igual que las *dis* puras.

#### 4. Conclusiones y Trabajo Futuro.

Este proyecto contribuye con la representación de una serie de patrones, algunos de los cuales pueden representarse con las herramientas conceptuales ya existentes en un *MER* convencional. Para los casos restantes, si es aplicable la heurística presentada en [12], se propone una transformación y en caso contrario se investigan formas alternativas de especificación.

Por lo anterior creemos que no es estrictamente necesario crear una nueva extensión del *MER*, al menos hasta no haber estudiado previamente en forma exhaustiva las posibilidades de representación de los conceptos aquí analizados con los mecanismos de abstracción ya estandarizados de facto. Adicionalmente, es necesario estudiar el efecto que la participación opcional introducen en las relaciones en presencia de las reglas del negocio, dado que se presentan diferentes situaciones cuando la participación en las relaciones no es total, dando lugar a relaciones redundantes en algunos casos, cuestión que altera la transformación.

Algunos autores han definido las dependencias de inclusión y exclusión teniendo en cuenta si los atributos que participan en las mismas provienen de relaciones o de entidades, si su dependencia es con atributos de una misma entidad o de otras entidades del modelo, o bien considerando si la dependencia se da con relación a conjuntos de tuplas de otras entidades o relaciones o solamente a instancias [7]. Teniendo en cuenta los resultados preliminares obtenidos, presumimos que estos conceptos quedarán embebidos en los casos catalogados, cuestión que se confirmará cuando se complete el conjunto de patrones de diseño.

#### 5. Referencias

- [1] Albrecht, M.; Buchholz, E.; Düsterhöft, A.; Thalheim, B. “*An Informal and Efficient Approach for Obtaining Semantic Constraints using Sample Data and Natural Language Processing*”, Lecture Notes In Computer Science, Vol. 1358, Selected Papers from a 1st. Workshop on Semantics in Databases, Prague, Czech Republic. Thalheim, B. and Libkin, L. (Eds.)1998.
- [2] Badía, A. “*Entity-Relationship Modeling Revisited*”, SIGMOD Record, Vol. 33, Nro. 1, 2004.
- [3] Chen, P. “*The Entity-relationship Model: Toward a Unified View of Data*” ACM Transactions on Database Systems. 1(1), 1976. Pag 9-36.
- [4] Chen, P. “*The Entity/Relationship Model: A bases for the enterprise view of data*” AFIPS Conference Proceedings, Vol. 46, 1977.
- [5] Codd, E. “*The Relational Model for Database Management*” Version 2. Addison Wesley Publ. Co, Reading, MA, 1990.
- [6] Date, C. “*An Introduction to Database Systems*”. Addison Wesley, Boston, MA, 2000.
- [7] De Miguel, A.; Piattini, M.; Marcos, E. “*Diseño de Bases de Datos Relacionales*”. Alfaomega – RaMa. 2000.
- [8] Elmasri, R.; Navathe, S. “*Fundamentals of Database Systems*”. Addison-Wesley, 3°Ed., 2000.
- [9] Halpin, T “*Conceptual Schema and Relational Database Design*”, 2°Ed.,1995.
- [10] Levene, M.; Loizou, G. “*A Guided Tour of Relational Databases and Beyond*” Springer-Verlag, London, 1999.
- [11] Rivero, L.; Doorn, J.; Ferraggine, V. “*Inclusion Dependencies*”. En Developing Quality Complex Database Systems: Practices, Techniques and Technologies, Shirley A. Becker(Editor). Idea Group Inc. Hershey(USA), 2001, pp. 261-278.
- [12] Rivero, L.; Ferraggine, V.; Doorn, J. “*Reingeniería de Bases de Datos: Un enfoque Basado en el Análisis de Dependencias de Inclusión*” Revista Colombiana de Computación. Vol. 4, 2003.
- [13] Rivero, L. “*An overview of Inclusión Dependencies*”. En Encyclopedia of Information Science and Technology. Khosrow-Pour M. (Editor). Idea Group. Inc. Hershey (USA). 2005.
- [14] Rochefeld, A y otros. “*MERISE, an Information System Design and Development Methodology*” 5<sup>th</sup> International Conference on Entity-Relationship Approach, Dijon, Francia, 1986.
- [15] Teorey, T.J. “*Database Modeling and Design: The Entity-Relationship Approach*”. Morgan Kaufmann Publishers, San Mateo, California, 1990.
- [16] UML [www.omg.org/uml](http://www.omg.org/uml)