

Las Ubicuas Dependencias de Comparación de Conjuntos de Valores: Conceptualización, Análisis y Modelado

Viviana E. Ferraggine¹ Laura C. Rivero^{1,2}

¹ INTIA, Facultad de Ciencias Exactas, U.N.C.P.B.A., Tandil, Buenos Aires, Argentina

² LINTI, Facultad de Informática, U.N.L.P. La Plata, Buenos Aires, Argentina
{vferra, lrivero} at exa.unicen.edu.ar

CONTEXTO

Este trabajo forma parte de las actividades científicas de la Línea "Integridad en Base de Datos guiada por Requisitos", del Proyecto "Base de Datos y Procesamiento de Señales". Este proyecto es continuación de proyectos previos en la temática y está actualmente en desarrollo en el Instituto de Investigación en Tecnología Informática Avanzada (INTIA) de la Facultad de Ciencias Exactas de la U.N.C.P.B.A.

El objetivo general de la línea es estudiar los problemas de la elicitación de conocimiento del Universo del Discurso (UdeD), así como el impacto de la realidad estudiada sobre los problemas de modelado e integridad en el contexto de bases de datos relacionales y post-relacionales.

RESUMEN

Como respuesta a la incesante necesidad de incorporar todo el conocimiento que encierran las reglas del negocio en los sistemas de información han surgido diversas propuestas para representar dichos conceptos en los modelos de datos desde los primeros estadios de su concepción. Sin embargo, la captura de estas reglas en el modelo conceptual no es una tarea simple, dado que son de naturaleza compleja, tienen una estructura arbitraria, son volátiles y difíciles de aplicar.

Dado que la existencia de conjuntos de atributos compatibles y semánticamente vinculados conduce frecuentemente a comparaciones entre ellos, éste trabajo se enfoca en un tipo especial de reglas del negocio que pueden representarse por medio de *Dependencias de Comparación de Conjuntos de Valores*. Ellas limitan el modo en el cual un conjunto de datos se relaciona con otro; es así que se pueden establecer cuatro modos: mediante relaciones de inclusión, de exclusión, de igualdad o de superposición parcial o solapamiento. Esta parte de nuestro proyecto está dedicada a estudiar dichas dependencias desde un punto de vista con centro en la estructura sintáctica de los atributos que están sujetos a comparación, haciendo uso del marco teórico que ha brindado previamente el análisis de las dependencias de inclusión.

Palabras clave: Diseño conceptual, dependencias de comparación de conjuntos de valores, dependencias de inclusión, dependencias de igualdad, dependencias de exclusión, dependencias de superposición parcial, reglas del negocio, modelo de datos.

1. INTRODUCCIÓN

El desarrollo de la tecnología de base de datos ha sido estimulado por la necesidad de garantizar la consistencia de los datos almacenados y la demanda de procesar datos más complejos con mayor contenido semántico. Estos factores han promovido un tratamiento más comprehensivo de las propiedades de los datos motivando la evolución de los Sistemas de Gestión de Base de Datos (Database Management Systems, DBMS) que paulatinamente han ido incorporando nuevas facilidades para manejar estos aspectos.

Se pueden identificar dos dificultades principales para manejar las propiedades de los datos dentro del contexto de un paradigma de base de datos: 1) Son difíciles de categorizar, en especial las que son evidentes en el mundo real; 2) Su representación varía durante el proceso de diseño de software. La primera cae en el campo de la Ingeniería de Requisitos [18]. La correcta elicitación de requisitos de un producto de software es un factor importante en el éxito del proceso de desarrollo de software pero no es fácil el tratamiento de estos requisitos. Ellos tienen diferente naturaleza intrínseca y

pueden aparecer mostrando distintos perfiles. En muchos casos, algunos requisitos están total o parcialmente ocultos en la información recogida por los diseñadores del software [17].

La segunda cae en el campo del Modelado Conceptual y Modelado Lógico Temprano o genérico. Durante estas actividades, las particularidades del modelo de datos y los paradigmas de bases de datos generan una diversidad de formas en las que ellos pueden ser considerados. Características pragmáticas tales como el grado de adhesión de un DBMS u otras relacionadas con el desempeño pueden introducir, como consecuencia, mutaciones adicionales.

No todas las propiedades de los datos deben ser modeladas. Ellas deben ser cuidadosamente estudiadas para elucidar si son necesarias dentro del alcance y el objetivo del artefacto de software.

Cuando una propiedad describe los valores permitidos de un dato, se incluye entre las Propiedades de los Datos. Otros aspectos semánticos describen cualidades de las asociaciones entre diferentes datos; éstas se conocen como Propiedades de las Conexiones entre Datos (o Relaciones). Cuando una propiedad de los datos conlleva una semántica que es específica del UdeD, se denomina Propiedad General. Este concepto también es conocido como regla de negocio, y es una declaración que define o restringe algunos aspectos del negocio. Éstas tienen la función de mantener la estructura comercial o controlar o influir en la conducta de negocio [8, 10, 22, 4]. En el mundo real, las propiedades de los datos se perciben desde un punto de vista semántico, permitiendo su agrupación según la complejidad semántica intrínseca y su clasificación de acuerdo a su esencia. Pero en el mundo de las bases de datos se perciben básicamente desde un punto de vista sintáctico.

A medida que el desarrollo del artefacto del software avanza, debe encontrarse una representación de estas propiedades según una estructura lícita del modelo. Esta representación será ciertamente diferente en el modelo de datos conceptual, en el modelo de datos lógico y en el modelo de datos del motor de la base de datos [19, 20, 21].

Como consecuencia de lo antedicho, han surgido variadas propuestas para representar estas reglas del UdeD en modelos de datos desde los primeros estadios de su concepción [22, 8, 4, 15], sin perder de vista que son generalmente complejas, tienen una estructura arbitraria y son volátiles.

Aunque los modelos conceptuales actuales han contribuido en gran medida a la captura de los aspectos relevantes del UdeD, no tienen el poder expresivo suficiente como para representar integralmente la diversidad natural de las reglas del negocio. Los medios formales o semiformales que ofrecen los modelos permiten representar parte de su semántica, sobre todo cuando se trata de reglas estáticas. Muchas de ellas pueden ser representadas en algunas de las variantes del Modelo de Entidades y Relaciones (MER) [9, 24], por ejemplo por medio de restricciones de cardinalidad y dependencias de existencia. Otras reglas que requieren la formulación explícita de eventos, condiciones y acciones específicos, no pueden ser representadas de un modo equivalente. Esta limitación deriva al menos en dos consecuencias. Por un lado, ha promovido la materialización de numerosas propuestas de ampliación a este modelo (devenido en el modelo semántico estándar "de facto" [2, 5, 6, 15, 16]) por ejemplo sugiriendo el uso de OCL (Object Constraint Language), de sintaxis ad-hoc y aún de lenguaje natural para enriquecer los diagramas representativos del esquema conceptual [13]. Por otro lado, algunas restricciones de integridad que podrían haberse representado adecuadamente con los mecanismos de abstracción provistos por el modelo, se difieren infundadamente hasta etapas más tardías del proceso de diseño y frecuentemente su materialización no utiliza los recursos expresivos más apropiados para el caso.

Desde otro punto de vista, existen enfoques en defensa del MER original proponiendo una rigurosa y amplia utilización de las capacidades que éste ofrece, sin las extensiones que dificultarían su uso, interpretación, validación y formalización [14]. Adhiriendo a este punto de vista, este proyecto se desarrolla en dos direcciones complementarias: 1) ofreciendo siempre que sea posible un modo de representar cada regla utilizando las construcciones básicas del MER (de no ser posible, se deberían diseñar patrones en lenguaje declarativo o en código de programación); 2) sugiriendo estrategias de transformación en función de lo obtenido en 1), cuando se ha plasmado una regla de forma incorrecta o ambigua.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Nuestras actividades de investigación están centradas actualmente en un tipo de reglas de sencilla interpretación en el mundo real pero con dificultades manifiestas de representación en el MER: las "*Dependencias de Comparación de Conjuntos de Valores*" (*dccv*), nombre genérico otorgado a las expresiones que permiten establecer comparaciones entre los valores de conjuntos de atributos estructuralmente compatibles y semánticamente relacionados. Esta clase de reglas incluye dependencias de inclusión (*di*) [7, 10]; dependencias de exclusión (*de*) [1, 12]; dependencias de igualdad (*dig*) y dependencias de superposición parcial (*dsp*) [23].

Las *dccv* se analizan desde una base sintáctica, a fin de establecer: a) la posibilidad de su representación en un MER utilizando las estructuras estandarizadas; b) la necesidad de ampliar el modelo para capturar la semántica de esas reglas, cuando las estructuras estandarizadas no sean suficientes, y c) la posibilidad de realizar transformaciones estructurales que podrían proporcionar algún beneficio en relación con la semántica o la calidad de los aspectos operativos, d) proponer otros medios de representación cuando los medios analíticos disponibles sean insuficientes, [19, 20, 21].

Formalmente se definen de la siguiente manera:

Sean R_i y R_d relaciones, denominadas término izquierdo y derecho respectivamente; W_i y W_d conjuntos de atributos compatibles de R_i y R_d . r_i y r_d indican las extensiones de R_i y R_d respectivamente. Siendo t una tupla, una subtupla de t correspondiente al conjunto de atributos W se indica como $t.W$. El Cuadro 1 contiene las definiciones de cada uno de los cuatro tipos de dependencias.

Una *dependencia de inclusión (di)* se define como la existencia de atributos en una relación cuyos valores deben ser un subconjunto de los valores de atributos compatibles en otra relación (o la misma). Formalmente una *di* es una expresión $R_i[W_i] \subseteq R_d[W_d]$ y se satisface si para $t \in r_i$ existe una tupla $t' \in r_d$, tal que $t.W_i = t'.W_d$. Si W_d es la clave primaria (K_d) de R_d , la *di* es basada en clave y usualmente se denomina restricción de integridad referencial (*rir*) o simplemente referencia. W_i es una clave extranjera (FK_i) de R_i , y la *rir* correspondiente se indica $R_i[FK_i] \ll R_d[K_d]$.

Una *dependencia de igualdad (dig)* se define como la existencia de un conjunto de atributos en una relación cuyos valores deben ser los mismos que los correspondientes a un conjunto de atributos compatibles en otra relación (o la misma). Esto significa que cada miembro del primer conjunto debe ser un miembro del segundo conjunto y viceversa. Formalmente una *dig* se indica $R_i[W_i] = R_d[W_d]$, esto se satisface si para $t \in r_i$, existe una tupla $t' \in r_d$ tal que $t.W_i = t'.W_d$ y $t' \in r_d$, existe una tupla $t \in r_i$ tal que $t.W_i = t'.W_d$. Una *dig* es equivalente a dos dependencias de inclusión inversas $R_i[W_i] \subseteq R_d[W_d] \wedge R_d[W_d] \subseteq R_i[W_i]$.

Una *dependencia de exclusión (de)* se define como la existencia de un conjunto de atributos en una relación cuyos valores deben ser mutuamente excluyentes respecto de los correspondientes a un conjunto de atributos compatibles en otra relación (o en la misma). Esto significa que ambos conjuntos de atributos no tienen miembros en común. Formalmente una *de* es una expresión $R_i[W_i] \parallel R_d[W_d]$, y se satisface si no existe $t \in r_d$ y $t' \in r_i$, tal que $t.W_d = t'.W_i$.

Por último, una *dependencia de superposición parcial (dsp)* se define como la existencia de un conjunto de atributos en una relación cuyos valores se solapan parcialmente con los correspondientes a un conjunto de atributos compatibles en otra relación (o en la misma). Se denotará $R_i[W_i] \cap R_d[W_d]$ y tiene un significado equivalente a: $R_d.W_d \not\subseteq R_i.W_i \wedge R_i.W_i \not\subseteq R_d.W_d \wedge R_d.W_d \not\parallel R_i.W_i$, siendo W_d y W_i conjuntos de atributos compatibles. Esto es $r_d.W_d \cap r_i.W_i \neq \emptyset \wedge r_d.W_d - r_i.W_i \neq \emptyset \wedge r_i.W_i - r_d.W_d \neq \emptyset$.

CUADRO 1: Tipos de Dependencias de Comparación de Conjuntos de Valores

En los últimos años la investigación de las *dis* ha categorizado muchos de sus aspectos [5, 6, 16]. Un marco de referencia para su estudio puede encontrarse en [10], como así también el estudio de su caso particular, las *rir*s [11, 19, 20, 21]. Respecto de las *dis*, un análisis de su posible origen desde una perspectiva semántica puede leerse en [19], y [9] es un estudio preliminar de las *dis*.

Para el análisis de las *dccvs* se ha seguido lo expuesto en [20, 21], examinando la estructura sintáctica de los conjuntos de atributos involucrados, independientemente del dominio de aplicación. Esto permite inferir el posible tipo de vínculo semántico que intenta plasmar la dependencia.

La estructura sintáctica de los términos de una dependencia puede definirse en función de la posición del término respecto de la clave de la relación (correlación). Siendo W un conjunto de atributos de una cierta relación R , K la clave primaria de R y Z un subconjunto de atributos no pertenecientes a la clave, la Figura 1 ilustra las cinco posiciones posibles de los términos.

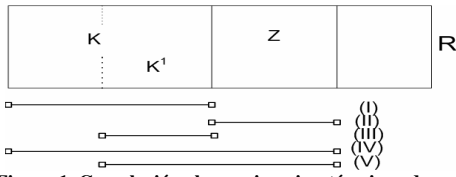


Figura 1. Correlación clave primaria - términos de una dependencia

I) $W \equiv K$; II) $W \equiv Z$; III) $W \equiv K^1$; IV) $W \equiv K \cup Z$ V) $W \equiv K^1 \cup Z$.
 K^1 es un subconjunto estricto de K , $K^1 \neq \emptyset$ y $Z \neq \emptyset$ para todos los casos. Teniendo en cuenta las correlaciones los 25 posibles pares de $\langle R_i, W_i, R_d, W_d \rangle$ quedan detallados en la Tabla 1.

Dado que las *dis* no son simétricas, la Tabla 1 muestra todas las posibles combinaciones de pares $\langle R_i, W_i, R_d, W_d \rangle$ y para las *des*, *digs* y *dsps* en las celdas sombreadas en gris se ha aclarado el caso para el cual resulta equivalente su análisis, por ejemplo el caso 2- $Z_i \theta K_d$ tiene un análisis equivalente (\equiv) al del caso 6- $K_i \theta Z_d$. θ es el tipo de *dcv*.

Tipo de Derecho	Término	Izquierdo	I	II	III	IV	V
		W_i	K_i	Z_i	K_i^1	$K_i \cup Z_i$	$K_i^1 \cup Z_i$
		W_d					
I	K_d		1- $K_i \theta K_d$	2- $Z_i \theta K_d$	3- $K_i^1 \theta K_d$	4- $K_i \cup Z_i \theta K_d$	5- $K_i^1 \cup Z_i \theta K_d$
II	Z_d		6- $K_i \theta Z_d$ ($\equiv 2$)	7- $Z_i \theta Z_d$	8- $K_i^1 \theta Z_d$	9- $K_i \cup Z_i \theta Z_d$	10- $K_i^1 \cup Z_i \theta Z_d$
III	K_d^1		11- $K_i \theta K_d^1$ ($\equiv 3$)	12- $Z_i \theta K_d^1$ ($\equiv 8$)	13- $K_i^1 \theta K_d^1$	14- $K_i \cup Z_i \theta K_d^1$	15- $K_i^1 \cup Z_i \theta K_d^1$
IV	$K_d \cup Z_d$		16- $K_i \theta K_d \cup Z_d$ ($\equiv 4$)	17- $Z_i \theta K_d \cup Z_d$ ($\equiv 9$)	18- $K_i^1 \theta K_d \cup Z_d$ ($\equiv 14$)	19- $K_i \cup Z_i \theta K_d \cup Z_d$	20- $K_i^1 \cup Z_i \theta K_d \cup Z_d$
V	$K_d^1 \cup Z_d$		21- $K_i \theta K_d^1 \cup Z_d$ ($\equiv 5$)	22- $Z_i \theta K_d^1 \cup Z_d$ ($\equiv 10$)	23- $K_i^1 \theta K_d^1 \cup Z_d$ ($\equiv 15$)	24- $K_i \cup Z_i \theta K_d^1 \cup Z_d$ ($\equiv 20$)	25- $K_i^1 \cup Z_i \theta K_d^1 \cup Z_d$

TABLA 1. Combinaciones de pares $\langle R_i, W_i, R_d, W_d \rangle$

3. RESULTADOS OBTENIDOS Y ESPERADOS

Siguiendo la primera dirección de desarrollo, se han determinado los patrones de representación de *dis* y *digs* en función de *rirs*, y se encuentra avanzado el mismo trabajo respecto de las *des* y *dsps*. En función de los logros anteriores, se han evaluado las ventajas y desventajas de aplicar las transformaciones según el nivel semántico del esquema original y la implementación de las operaciones de actualización básicas. Se pueden resumir las siguientes conclusiones parciales:

- Las operaciones de actualización que es factible llevar a cabo en el esquema original, también son factibles en el esquema transformado. Del mismo modo las inaplicables en uno y otro esquema son las mismas.
- Varias de las transformaciones estudiadas generan caminos cíclicos de integridad referencial que requieren una especificación por medio de aserciones o código ad hoc, siendo el "rechazo" la acción en caso de violación.
- Existen casos que pueden resolverse aplicando algunos axiomas de transformación básicos del modelo relacional (específicamente regla de pullback y axioma de transitividad). Otros mejoran la calidad de la representación mediante sencillas transformaciones fundamentadas en atributos y relaciones inaplicables. Se ha determinado el conjunto de casos para los cuales la calidad semántica no mejora porque las transformaciones aún siendo aplicables no permiten revelar objetos y relaciones mal representados, o bien porque no hay transformación que sea aplicable.
- Siempre que sea conveniente la transformación, el nuevo esquema adhiere al MER convencional o al MER extendido con agregaciones.
- Con respecto a razones de diseño, los esquemas transformados generalmente introducen nuevas tablas. Esto introduce un nuevo contexto de desempeño que debería ser analizado teniendo en cuenta las características del negocio (estabilidad de la información, frecuencia de operaciones de actualización, etc.)

- Finalmente, el esquema resultante debe re-analizarse basado en las expresiones de multiplicidades asociadas a los roles de las entidades participantes para determinar la validez de los caminos cíclicos y eliminar relaciones redundantes. Debido al agregado de tablas y relaciones se hace necesario un nuevo análisis de la coherencia global del esquema resultante. El procedimiento de validación del esquema conceptual permite la detección de multiplicidades y participaciones incorrectas [3].

La aplicación de transformaciones (en conjunción con el análisis de vínculos semánticos redundantes y el uso de mecanismos de evaluación de esquemas) ha resultado un marco metodológico apropiado para lograr una parcial pero útil reingeniería de esquemas de base de datos relacionales de pobre calidad, desactualizados o erróneos. La fase de transformación contribuye en el mejoramiento del nivel semántico del esquema, poniendo en evidencia objetos mal modelados y relaciones que los unen con otros objetos. Las transformaciones aplicables a las dependencias de inclusión y concretadas a través de una heurística fueron apropiadas también para las dependencias de igualdad. Algunos casos no pudieron mejorarse con este enfoque debido a que persisten dependencias intratables. No es posible establecer una regla estricta a seguir, pero generalmente, si una relación es conceptualmente relevante o las operaciones de actualización son poco frecuentes, la transformación es conveniente aunque aumentará el número de entidades y relaciones. Por otro lado, cuando el esquema resultante comience a ser poblado por entidades y relaciones antinaturales, el procedimiento de transformación debería detenerse.

Las *des* y *dsps* serán enfocadas desde otra perspectiva ya que ninguna de ambas está en el espíritu del MER, con la excepción de las jerarquías exclusivas y compartidas respectivamente.

Dentro de los futuros trabajos de investigación se incluye además el análisis de las acciones que podrían desencadenar las *dccvs* haciendo una analogía con las acciones referenciales de las *rirs* y el estudio de patrones de código (triggers) para soportar las *dis* remanentes en cada caso.

4. FORMACION DE RECURSOS HUMANOS

La investigación desarrollada ha permitido elaborar un programa para materias optativas centradas en tópicos avanzados de modelado conceptual de datos. En el mismo sentido varios grupos de alumnos han desarrollado o han profundizado en aspectos conceptuales relativos a este tema durante el desarrollo de su tesis de grado para aspirar al título de Ingeniero de Sistemas-UNCPBA.

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, 1358, pp. 1-28. Selected Papers from a 1st. Workshop on Semantics in Databases, Czech Republic. Thalheim, B.&Libkin, L.(Eds.), 1998.
- [2] Badía, A., "Entity-Relationship Modeling Revisited", SIGMOD Record, 33(1), 2004.
- [3] Boufarés, F., Bennaceur, H., "Consistency Problems in ER-Schemas for Database Systems", Information Sciences 163, pp. 263-274, 2004.
- [4] Business Rule Group, "Defining Business Rules ~ What Are They Really?", 1995, http://www.businessrulesgroup.org/first_paper/br01c0.htm.
- [5] Camps, R., "From Ternary Relationship to Relational Tables: a Case Against Common Beliefs", SIGMOD Record, 31(2), pp. 46-49, 2002.
- [6] Camps, R., "Transforming N-Ary Relationships to Database Schemas: an Old and Forgotten Problem", RR LSI-02-5-R, Univ.Politécnica de Catalunya, Spain, 2002.
- [7] Casanova, M., Fagin, R., Papadimitriou, C., "Inclusion Dependencies and their Interaction with Functional Dependencies", J. Computer and System Sciences, 28(1), 1984.
- [8] Ceri, S., Fraternali, P., "Designing Database Applications with Objects and Rules. The IDEA Methodology", Addison Wesley Publ. Co, 1997.
- [9] Chen, P., "The Entity-relationship Model: Toward a Unified View of Data", ACM TODS, 1(1), pp. 9-36, 1976.
- [10] Codd, E., "The Relational Model for Database Management. Version 2", Addison Wesley Publ. Co., 1990.
- [11] Cuadra, D., Nieto, C., Martínez, P., Castro, E., Velasco, M., "Preserving Relationship Cardinality Constraints in Relational Schemata", In Database Integrity: Challenges and Solutions, Idea Group Inc., Jorge Doorn and Laura Rivero Editors, pp. 66-112, 2002.

- [12] De Miguel, A., Piattini, M., Marcos, E., "Diseño de Bases de Datos Relacionales", Alfaomega Grupo Editor, 2000.
- [13] Goelman, D., Hussmann, H., "Using UML/OCL Constraints for Relational Database Design", In UML'99: The Unified Modeling Language - Beyond the Standard, Lecture Notes in Computer Science, Springer, 1723, France, R. & Rumpe, B. Editors, pp. 598-613, 1999.
- [14] Goelman, D., Song, I-Y., "Entity-Relationship Modeling Re-revisited", ER 2004, Proceedings 23rd International Conference on Conceptual Modeling, Shanghai, China. Lecture Notes in Computer Science, Springer, 3288, Paolo Atzeni, Wesley W. Chu, Hongjun Lu, Shuigeng Zhou, Tok Wang Ling Editors, pp. 43-54, 2004.
- [15] Halpin, T., "Information Modeling and Relational Databases. From Conceptual Analysis to Logical Design", Morgan Kaufmann Publishers, 2001.
- [16] Jones, T., Song, I-Y., "Analysis of Binary/Ternary Cardinality Combinations in Entity-Relationship Modeling", Data & Knowledge Engineering, 19 (1), pp. 39-64, 1996.
- [17] Jackson, M. (1995). Software Requirements & Specifications. A Lexicon of Practice, principles and prejudices. Workingham: Addison Wesley Publishers, ACM Press.
- [18] Loucopoulos, P & Karakostas. (1995). System Requirements Engineering. McGraw Hill International Series in Software Engineering.
- [19] Rivero, L., "Inclusion Dependencies", In Encyclopedia of Information Science and Technology. Idea Group. Inc., Mehdi Khosrow-Pour Editor, pp. 1425-1430, 2005.
- [20] Rivero, L., Doorn, J., Ferraggine, V., "Inclusion Dependencies". In Developing Quality Complex Database Systems: Practices, Techniques and Technologies, Idea Group Inc., Shirley A. Becker Editor, pp 261-278, 2001.
- [21] Rivero, L., Doorn, J., Ferraggine, V., "Enhancing Relational Schemas through the Analysis of Inclusion Dependencies", Int'l Journal of Computer Research, 12(4), Nova Publishers, pp. 489-511, 2004.
- [22] Ross, Ronald G., "The Business Rule Book. Classifying, Defining and Modeling Rules", Database Research Group, R. Ross Editor/Publisher, 1997.
- [23] Tari, Z., Bukhres O., Stokes J., Hammoudi S., "The Reengineering of Relational Databases Based on Key and Data Correlations", In Searching for Semantics: Datamining, Reverse Engineering, etc. Proc.7th IFIP 2.6 Working Conference on Database Semantics (DS-7), Chapman & Hall, Lausanne, S. Scappapietra and F. Maryanski Editors, pp. 183-214, 1998.
- [24] Teorey, T.J., "Database Modeling and Design. The Entity-Relationship Approach", Morgan Kaufmann Publishers, 1990.