

IV Jornadas de
Ingeniería del Software
y Bases de Datos
(JISBD'99)



Editores:
Pere Botella
Juan Hernández
Félix Saltor

Grupo de Paralelismo
Departamento de Informática
Escuela Politécnica
Universidad de Extremadura

Cáceres,
del 24 al 26 de noviembre de 1999

IV Jornadas de Ingeniería del Software y Bases de Datos (JISBD'99)

Editores:

Pere Botella
Juan Hernández
Fèlix Saltor

Grupo de Paralelismo
Departamento de Informática
Escuela Politécnica
Universidad de Extremadura

Cáceres, del 24 al 26 de noviembre de 1999

ACTAS DE LAS IV JORNADAS DE INGENIERÍA DEL SOFTWARE
Y BASES DE DATOS (JISBD'99)

EDITORES:

Pere Botella
Juan Hernández
Fèlix Saltor

ORGANIZADAS POR:

Grupo de Paralelismo
Departamento de Informática
Escuela Politécnica
Universidad de Extremadura

ENTIDADES COLABORADORAS:

ATI Asociación de Técnicos de Informática
Caja Extremadura
Diputación Provincial de Cáceres. Institución
Cultural EL BROCENSE
Dirección General de Enseñanza Superior e
Investigación Científica
Junta de Extremadura. Consejería de Educación,
Ciencia y Tecnología
Teleserver Extremadura

© Los autores

Primera edición, 1999

I.S.B.N.: 84-699-0956-8

Depósito legal: CC-238-1999

Imprime: Gráficas Morgado, S.L. - Cáceres.

Aldana, J.F.
Alvarez, D.
Aramburu,
Bardanga, F.
Blesa, P.
Basabea, N.
Calero, C.
Cerveron,
Corral, A.
Cuadra, D.
Cueva, J.M.
De Miguel,
Delgado, C.
Diaz, O.
Duran, A.
Duran, M.
Garcia, F.
Garcia, S.
Gavi, E.
Gomez, A.
Gomis, O.
Jaime, A.
Kaufman,
Lain, C.
Lopez, J.
Lopez, A.
Manzana,
Marcos,
Martin, C.
Martinez,
Martinez,
May, A.
Miguel, L.
Moreno,
Nieto, C.
Ortega,
Ortin, F.
Parama,
Pazos, S.
Penaba,
Perez, J.
Perez, J.
Pillipi,
Places,
Polo, M.
Polo, A.
Ruiz, E.
Saez,
Safis,
Samos,
Sanz, J.
Talens,
Toro, I.
Torres,
Vassil,
Vela, I.

COMITÉ DE PROGRAMA DE BASES DE DATOS

PRESIDENTE

Fèlix Saltor (Universitat Politècnica de Catalunya)

MIEMBROS

José Francisco Aldana	(Universidad de Málaga)
M ^a . José Aramburu	(Universitat Jaume I)
Manuel Barrena	(Universidad de Extremadura)
Pedro Blesa	(Universitat Politècnica de València)
Nieves R. Brisaboa	(Universidade da Coruña)
José M ^a . Busta	(Universidade de Santiago de Compostela)
Matilde Celma	(Universitat Politècnica de València)
Adoración de Miguel	(Universidad Carlos III de Madrid)
Jesús García Molina	(Universidad de Murcia)
Alfredo Goñi	(Euskal Herriko Unibertsitatea)
Arantza Illarramendi	(Euskal Herriko Unibertsitatea)
Esperanza Marcos	(Universidad Rey Juan Carlos)
Eduardo Mena	(Universidad de Zaragoza)
Roberto Moriyón	(Universidad Autónoma de Madrid)
Mario Piattini	(Universidad de Castilla-La Mancha)
Antonio Polo	(Universidad de Extremadura)
José Samos	(Universidad de Granada)
Toni Urpí	(Universitat Politècnica de Catalunya)

COLABORADORES EN EL PROCESO DE REVISIÓN

Rafael Berlanga	Marcela F. Genero
Coral Calero	Margarita Martínez
Dolors Costal	Miryam Salas
Cecilia Delgado	

PRÓLOGO

El presente volumen contiene los trabajos presentados en las IV Jornadas de Ingeniería de Software y las IV Jornadas de Bases de Datos, *JISBD'99*, celebradas en Cáceres los días 24, 25 y 26 de noviembre de 1999.

Tanto las Jornadas de Ingeniería de Software como las Jornadas de Investigación y Docencia en Bases de Datos han constituido, en los últimos años, la referencia obligada para el intercambio de experiencias de los grupos españoles que desarrollamos nuestra investigación en estos campos. Asimismo, hay que destacar el notable incremento en la participación de grupos de investigación Iberoamericanos en esta cuarta edición.

Las IV Jornadas de Ingeniería de Software y Bases de Datos representan una edición especial, donde coinciden por primera vez estos dos eventos nacionales ya consolidados: de una parte, las Jornadas de Ingeniería de Software, que tuvieron lugar en Sevilla, San Sebastián y Murcia en los años 1996, 1997 y 1998, respectivamente; de otra, las Jornadas de Investigación y Docencia en Bases de Datos, celebradas con anterioridad en A Coruña, Madrid y Valencia en esos mismos años.

El objetivo principal de estas Jornadas es propiciar un foro de encuentro y discusión donde los grupos de investigación en las áreas de Ingeniería de Software y Bases de Datos, tanto universitarios como profesionales, puedan compartir experiencias, divulgar resultados, y donde se establezca un marco propicio de colaboración entre los distintos sectores.

Las IV Jornadas se han organizado en doce sesiones técnicas paralelas, dos conferencias invitadas y dos mesas redondas.

Para las sesiones técnicas se recibieron un total de 61 trabajos en Ingeniería del Software y 26 trabajos en Bases de Datos. Cada artículo fue revisado por al menos dos miembros del Comité de Programa, quienes tras una ardua labor dado el número y calidad de las contribuciones recibidas, seleccionaron 23 trabajos en el área de Ingeniería de Software y 11 en el área de Bases de Datos. Adicionalmente, en cada una de las áreas se consideraron 8 y 5 trabajos para su presentación en la modalidad de artículos cortos, respectivamente. A mediados del mes de julio, los autores recibieron los comentarios y opiniones de los miembros de los respectivos comités relativos a sus trabajos, para incorporar en las versiones finales recogidas en este volumen, las modificaciones recomendadas por los revisores.

Las conferencias invitadas se han organizado teniendo en cuenta las inquietudes e intereses de ambos colectivos, y tuvimos el honor de contar con la presencia de personalidades prestigiosas en sus respectivas especialidades. El Dr. Mehmet Aksit, profesor de la Universidad de Twente (Holanda), director y líder del grupo TRESE (*Twente Research and Education on Software Engineering*), quien impartió la conferencia "Quality-Oriented Software Engineering". El Dr. Stephen Cannan, de *james martin + co* (Holanda) y miembro del *ISO Database Languages Working Group*, quien impartió la conferencia "The New SQL standard. Good, Bad or simply Ugly?"

Las mesas redondas trataron temas de interés comunes a ambos colectivos. De un lado "¿Para qué la tesis doctoral?", coordinada por el Dr. Oscar Díaz. De otro, "La Ingeniería de Software y las Bases de Datos en la Industria: cómo trasladar la investigación a la práctica", coordinada por el Dr. Mario Piattini. A todos los participantes en las mismas, quisiera desde aquí hacer constar también mi agradecimiento.

Estas jornadas no hubieran sido posibles sin el trabajo voluntario y desinteresado de muchas personas. A todos los miembros de los respectivos Comités de Programa y a los revisores externos, quisiera agradecerles el esfuerzo realizado en todo el proceso de revisión y selección de los trabajos. A sus dos Presidentes, Dr. Pere Botella y Dr. Fèlix Saltor, de la Universitat Politècnica de Catalunya, mi gratitud por el esfuerzo realizado y la colaboración prestada en todo momento.

Asimismo, quiero agradecer a todos los miembros del Comité Organizador el buen ánimo que han tenido y las horas empleadas que, a veces, sin tiempo disponible debido a otras ocupaciones no menos importantes, han dedicado en la realización de la multitud de tareas que conlleva la organización de este tipo de eventos.

Finalmente y en nombre de todo el Comité Organizador, quiero expresar nuestra gratitud a todos los autores que enviaron sus artículos y a todos los participantes en estas Jornadas, sin los cuales, esta cuarta edición no hubieran tenido lugar. Todo el Comité Organizador espera que este encuentro haya sido de vuestro interés.

Cáceres, noviembre de 1999

Juan Hernández

CONTENIDOS

Conferencias invitadas

Quality-Oriented Software Engineering
Dr. Mehmet Aksit (Universidad de Twente, Holanda)

The New SQL Standard: Good, Bad, or Simply Ugly?
Stephen Cannan (james martin+co, Holanda)
 (Convenor - ISO Database Languages Working Group)

Mesas redondas

¿Para qué la tesis doctoral? 1
Moderador: Oscar Díaz (Euskal Herriko Unibertsitatea)

La Ingeniería del Software y las Bases de Datos en la
 Industria: cómo trasladar la investigación a la práctica 7
Moderador: Mario Piattini (Universidad de Castilla-La Mancha)

Sesiones técnicas de Bases de Datos

BD1: Bases de datos orientadas a objetos 11
Moderador: Jesús García-Molina (Universidad de Murcia)

Definición de clases derivadas y clases con aumento de la
 capacidad de información en bases de datos orientadas a objetos 13
J. Samos, C. Delgado, E. Garvía, E. Ruiz

Propuesta de una extensión del ODMG para el soporte de
 relaciones ternarias 25
E. Marcos, B. Vela, J. Sáenz

Diseño de un sistema de persistencia implícita mediante reflectividad computacional	39
<i>F. Ortín, A.B. Martínez, D. Álvarez, J.M. Cueva</i>	
BD2: Artículos cortos	51
<i>Moderador: Toni Urpí (Universitat Politècnica de Catalunya)</i>	
Profundizando en la semántica de las cardinalidades en el modelo E/R extendido	53
<i>P. Martínez, C. Nieto, D. Cuadra, A. de Miguel</i>	
Semántica formal de asociaciones y agregados para su implementación con reglas activas	55
<i>O. Martín, J. Torres, A. Durán, J.A. Pérez, M. Toro</i>	
Prestaciones de la copia primaria en bases de datos distribuidas: estudio comparativo	57
<i>P. Blesa, R. Cerverón, O. Gomis, S. Talens</i>	
Interfaz JDBC para un sistema de administración de bases de datos distribuidas	59
<i>R. Ortega, R. Pazos, A. May, J. Pérez</i>	
Esquemas de indexación espacio-temporales para datos microclimáticos ..	61
<i>A. Corral, M. Vassilakopoulos</i>	
BD3: Bases de datos distribuidas y heterogéneas	63
<i>Moderador: Eduardo Mena (Universidad de Zaragoza)</i>	
Un sistema de meta-base de datos basado en UML e integrado en un sistema de gestión de bases de datos distribuidas	65
<i>M.N. Moreno, M.J. Polo, L.A. Miguel, F.J. García</i>	
Prototipo para la detección y registro de consultas relevantes en bases de datos distribuidas	75
<i>J. Pérez, R. Kauffmann, R. Pazos, S. Martínez</i>	
Propuesta de arquitectura para un sistema de bases de datos documentales	85
<i>N.R. Brisaboa, M.J. Durán, C. Lalín, J.R. López, J.R. Paramá, M.R. Penabad, Á.S. Places</i>	

BD4: Métodos de acceso e indexación	97
<i>Moderador:</i> Manuel Barrena (Universidad de Extremadura)	
Avances en el diseño del índice m-Q paginado	99
<i>M. Salas, A. Polo, A.I. López, J.C. Manzano</i>	
Chronology: Una aproximación al almacenamiento y recuperación de información de actualidad	111
<i>R. Berlanga, M.J. Aramburu, S. García, I. Sanz</i>	
Revisando las técnicas de optimización de consultas recursivas mediante la propagación de semicomposiciones	123
<i>J.F. Aldana, A.C. Gómez</i>	
BD5: El colectivo de investigadores de bases de datos y sus inquietudes: su estructuración, investigación y docencia	135
<i>Moderadores:</i> Adoración de Miguel (Universidad Carlos III de Madrid) Fèlix Saltor (Universitat Politècnica de Catalunya)	
BD6: Diseño y depuración de bases de datos	137
<i>Moderadora:</i> Matilde Celma (Universitat Politècnica de València)	
Caracterización formal de métricas para bases de datos relacionales	139
<i>C. Calero, M. Piattini</i>	
Depuración de disparadores en bases de datos activas	151
<i>O. Díaz, A. Jaime</i>	

Sesiones técnicas de Ingeniería del Software

ISO: Artículos cortos	163
<i>Moderador:</i> Pere Botella (Universitat Politècnica de Catalunya)	
Especialización en el ámbito del modelado conceptual	165
<i>P. Sánchez, F. Letelier, J. A. Troyano, V. Pelechano, J. Torres</i>	
Modelado orientado a objetos basado en UML conectando a una herramienta de diseño detallado	167
<i>M. López, I.A. Parra, R. Santaolaya, O.G. Frago, B.J. Ferro</i>	

CARACTERIZACIÓN FORMAL DE MÉTRICAS PARA BASES DE DATOS RELACIONALES¹

Coral Calero y Mario Piattini
Grupo ALARCOS
Dpto. Informática. U.C.L.M.
e-mail: {ccalero, mpiattin}@inf-cr.uclm.es

Resumen. Las bases de datos relacionales se han convertido en uno de los componentes más importantes de los Sistemas de Información (SI) aunque su diseño sigue siendo una tarea larga, difícil y costosa. Las metodologías existentes sólo recomiendan utilizar la teoría de la normalización para comprobar la calidad de los esquemas relacionales, sin embargo, consideramos fundamental el desarrollo de métricas específicas para bases de datos relacionales que nos permitan asegurar su calidad. Dentro de los factores de calidad destaca por su importancia la mantenibilidad, que se ve influida por la complejidad. En este artículo se proponen diferentes métricas para medir la complejidad de bases de datos relacionales, caracterizándolas dentro del marco formal de Zuse [30].

1 Introducción

En los últimos años hemos asistido a la extraordinaria difusión de las bases de datos relacionales, fuertemente implantadas desde en los grandes ordenadores hasta en los ordenadores personales. Las bases de datos y, especialmente las relacionales, se han convertido en el corazón de los Sistemas de Información más relevantes para el funcionamiento de la sociedad. Sin embargo, su diseño sigue siendo una tarea larga, difícil y costosa [9], a pesar de las ventajas que se derivan del sólido fundamento matemático desarrollado desde la definición del modelo relacional [6].

Como sabemos, la mayor parte de las metodologías existentes [2], [17], [7], [10], [23] propone abordar el diseño de bases de datos relacionales en tres grandes etapas: modelado conceptual, diseño lógico y diseño físico; aplicando una serie de reglas que permiten transformar esquemas conceptuales en esquemas relacionales, y comprobando la "calidad" de los esquemas relacionales obtenidos utilizando la teoría de la normalización. A pesar de lo útil que puede resultar la teoría de la normalización, es necesario elaborar nuevas métricas para evaluar la calidad de una base de datos relacional. De hecho, actualmente se reconoce que la medición del software es una forma eficaz de entender, monitorizar, controlar, predecir y mejorar los proyectos de desarrollos y mantenimiento del software [3]. Además, puede ayudar a los profesionales e investigadores a tomar las decisiones más adecuadas [25], guiando al diseñador a la hora de elegir entre esquemas de bases de datos alternativos.

¹ Este trabajo forma parte del proyecto MANTICA, parcialmente financiado por la CICYT y la Unión Europea (1FD97-0168)

Desafortunadamente, desde que McCabe propusiera la complejidad ciclomática [22] hasta nuestros días, la mayor parte de las métricas se ha centrado en las características de los programas dejando totalmente a un lado las bases de datos [27]. Sin embargo, resulta fundamental elaborar métricas de diseño para bases de datos ya que su tamaño y naturaleza pueden influir, en gran medida, en muchos aspectos de un SI como el esfuerzo de desarrollo [21].

Uno de los aspectos más importantes relativos a la calidad [19], es la mantenibilidad debido, básicamente, a los recursos humanos y materiales que comporta. La mantenibilidad depende de tres factores: entendibilidad, modificabilidad y facilidad de prueba, los cuales a su vez están influidos por la complejidad [20]. Sin embargo, conseguir una medida general para la complejidad de una base de datos relacional es como perseguir el santo grial [13]. Siguiendo a [18] la complejidad puede dividirse en: complejidad representacional, computacional y psicológica. Esta última, a su vez, puede considerarse compuesta por: la complejidad del problema, los factores cognitivos humanos y la complejidad del producto, en la que se centra actualmente nuestra investigación.

En este artículo proponemos un conjunto de métricas para medir la complejidad de esquemas de bases de datos relacionales: NA (Número de atributos), DRT (Longitud del camino referencial) y RD (Grado de referenciabilidad), caracterizándolas formalmente dentro del marco propuesto por [30]. En el apartado 2 presentamos las métricas, pasando en el tercero a hacer una pequeña introducción del marco formal utilizado para, en el apartado siguiente, presentar la verificación formal de las métricas. Por último, exponemos las conclusiones y líneas futuras de trabajo.

2 Métricas para la complejidad de bases de datos relacionales

Desde que a finales de los sesenta el Dr. Codd propusiera su modelo relacional [6], se ha intensificado la investigación en el campo de las bases de datos y los productos de bases de datos relacionales han generado una importante industria. [8] define un sistema de gestión de base de datos relacional como "un sistema, en el que como mínimo:

- Los datos son vistos por el usuario como tablas (y sólo así) y
- Los operadores disponibles para el usuario generan nuevas tablas a partir de otras antiguas. Entre los operadores se encuentran, como mínimo la selección, (SELECT), la proyección (PROJECT) y la combinación (JOIN)".

El único indicador utilizado para medir la calidad de una base de datos relacional ha sido la teoría de la normalización, a partir de la cual [16] proponen un ratio de normalidad.

En este artículo, proponemos las siguientes métricas para medir la complejidad de una base de datos relacional:

- **Número de atributos** (*Number of Attributes NA*). $NA(A)$ es el número de atributos de la tabla A.
- **Grado de Referenciabilidad** (*Referentiability Degree RD*). $RD(A)$ es el número de claves ajenas de la tabla A.

•Profundidad del árbol referencial (*Depth Referential Tree DRT*). $DRT(A)$ se define como la longitud del máximo camino referencial de la tabla A. Los ciclos sólo se consideran una vez. Aplicamos las métricas definidas al ejemplo de la tabla 1, tomado de [12]:

```

CREATE TABLE EMPLEADO
(
  NOMBREP VARCHAR(15) NOT NULL,
  INIC CHAR,
  APELLIDO VARCHAR(15) NOT NULL,
  NSS CHAR(9) NOT NULL,
  FECHAEN DATE,
  DIRECCION VARCHAR(30),
  SEXO CHAR,
  SALARIO DECIMAL(10,2),
  NSSSUPER CHAR(9),
  ND INT NOT NULL,
  CONSTRAINT CLPEMP
    PRIMARY KEY (NSS),
  CONSTRAINT CLESUPEREMP
  FOREIGN KEY (NSSSUPER) REFERENCES EMPLEADO(NSS)
  ON DELETE SET NULL ON UPDATE CASCADE,
  CONSTRAINT CLEDEPTOEMP
  FOREIGN KEY (ND) REFERENCES
    DEPARTAMENTO (NUMEROD)
  ON DELETE SET DEFAULT ON UPDATE CASCADE);

CREATE TABLE DEPARTAMENTO
(
  NOMBRED VARCHAR(15) NOT NULL,
  NUMEROD INT NOT NULL,
  NSSGTE CHAR(9) NOT NULL,
  FECHAINICGTE DATE,
  CONSTRAINT CLPDEPTO
    PRIMARY KEY (NUMEROD),
  CONSTRAINT CLSDEPTO
    UNIQUE(NOMBRED),
  CONSTRAINT CLEGTDEPTO
  FOREIGN KEY (NSSGTE) REFERENCES EMPLEADO(NSS)
  ON DELETE SET DEFAULT ON UPDATE
  CASCADE);

CREATE TABLE LUGAR_DEPTS
(
  NUMEROD INT NOT NULL,
  LUGARD VARCHAR(15) NOT NULL,
  PRIMARY KEY (NUMEROD, LUGARD),
  FOREIGN KEY (NUMEROD) REFERENCES
    DEPARTAMENTO (NUMEROD)
  ON DELETE CASCADE ON UPDATE CASCADE);

CREATE TABLE PROYECTO
(
  NOMBREPR VARCHAR(15) NOT NULL,
  NUMEROPR INT NOT NULL,
  LUGARPR VARCHAR(15),
  NUMD INT NOT NULL,
  PRIMARY KEY (NUMEROPR),
  UNIQUE (NOMBREPR),
  FOREIGN KEY (NUMD) REFERENCES
    DEPARTAMENTO (NUMEROD));

CREATE TABLE TRABAJA_EN
(
  NSSE CHAR(9) NOT NULL,
  NUMP INT NOT NULL,
  HORAS DECIMAL(3,1) NOT NULL,
  PRIMARY KEY (NSSE, NUMP),
  FOREIGN KEY (NSSE) REFERENCES EMPLEADO
  (NSS),
  FOREIGN KEY (NUMP) REFERENCES
    PROYECTO (NUMEROPR));

CREATE TABLE DEPENDIENTE
(
  NSSE CHAR(9) NOT NULL,
  NOMBRE_DEPEND VARCHAR(15)
  NOTNULL,
  SEXO CHAR,
  FECHAAN DATE,
  RELACION VARCHAR(8),
  PRIMARY KEY (NSSE, NOMBRE_DEPEND),
  FOREIGN KEY (NSSE) REFERENCES EMPLEADO
  (NSS));
    
```

Tabla 1. Ejemplo

En la figura 1 se presenta el grafo relacional para el ejemplo anterior, donde las flechas indican las relaciones de integridad referencial.

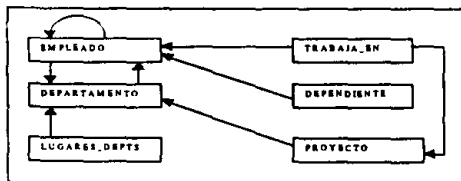


Figura 1. Esquema del ejemplo

	NA	RD	DRT
EMPLEADO	10	2	3
DEPARTAMENTO	4	1	3
LUGARES DEPTS	2	1	4
PROYECTO	4	1	4
TRABAJA EN	3	2	5
DEPENDIENTE	5	1	4
ESQUEMA	28	8	5

Tabla 2. Valores de las métricas para el ejemplo

Los valores de las métricas correspondientes al ejemplo se muestran en la tabla 2, donde podemos observar todos los valores obtenidos para las métricas a partir de la información que nos da tanto el ejemplo como su esquema. Por ejemplo, el valor para la métrica DRT para la tabla TRABAJA_EN es cinco, que se obtiene del camino referencial siguiente (conviene recordar que, por definición, nuestra métrica DRT sólo considera los ciclos una vez): TRABAJA_EN → PROYECTO → DEPARTAMENTO → EMPLEADO → EMPLEADO → DEPARTAMENTO.

3 Introducción al marco formal de Zuse [30]

Varios han sido los autores que han propuesto marcos para caracterizar las métricas [3], [14], [29]. En este artículo, seguimos el marco formal propuesto por [30] para describir las propiedades de las métricas propuestas anteriormente. Este marco es una extensión de la teoría de la medida clásica, en el que se da tanto una base para validar las métricas del software como criterios sobre las escalas de las medidas. Zuse describe la medida como algo "necesario ya que los humanos no son capaces de tomar decisiones o enjuiciar de forma clara y objetiva". Medir es más que producir números, es la combinación de entidades empíricas con entidades numéricas. El interés se centra en establecer "relaciones empíricas" entre objetos, del tipo "mayor que" o "mayor igual que". Estas relaciones empíricas se representan mediante los símbolos " $\bullet >$ " y " $\bullet \geq$ " respectivamente.

Un Sistema Relacional Empírico se representa como: $A = (A, \bullet \geq, \circ)$, donde A es un conjunto no vacío de objetos, $\bullet \geq$ es una relación empírica en A y \circ es una operación binaria cerrada (concatenación) en A . Mientras que un Sistema Numérico Relacional puede ser definido como: $B = (\mathfrak{R}, \geq, +)$, donde \mathfrak{R} son los números reales, \geq una relación en \mathfrak{R} , y $+$ una operación binaria en \mathfrak{R} .

Por tanto una medida va a ser un correspondencia $u: A \rightarrow \mathfrak{R}$ tal que para todo a, b pertenecientes al mundo real A se cumple que: $a \bullet \geq b \Leftrightarrow u(a) \geq u(b)$. El autor define un conjunto de propiedades para las medidas las cuales caracterizan diferentes estructuras de medida. Las más importantes se presentan en la tabla 3:

Estructura Extensiva Modificada	Condiciones de Independencia	Relación Modificada de Creencia (<i>belief</i>)
Axioma1: $(A, \bullet \geq)$ (orden débil)	C1: $A1 = A2 \Rightarrow A1 \circ A = A2 \circ A$ y $A1 = A2 \Rightarrow A \circ A1 = A \circ A2$	MRB1: $\forall A, B \in \mathfrak{S}: A \bullet \geq B \circ B \bullet \geq A$ (compleción)
Axioma2: $A1 \circ A2 \bullet \geq A1$ (positividad)	C2: $A1 = A2 \Leftrightarrow A1 \circ A = A2 \circ A$ y $A1 = A2 \Leftrightarrow A \circ A1 = A \circ A2$	MRB2: $\forall A, B, C \in \mathfrak{S}: A \bullet \geq B$ y $B \bullet \geq C \Rightarrow A \bullet \geq C$ (transitividad)
Axioma3: $A1 \circ (A2 \circ A3) = (A1 \circ A2) \circ A3$ (asociatividad débil)	C3: $A1 \bullet \geq A2 \Rightarrow A1 \circ A \bullet \geq A2 \circ A$, y $A1 \bullet \geq A2 \Rightarrow A \circ A1 \bullet \geq A \circ A2$	MRB3: $\forall A \supseteq B = A \bullet \geq B$ (axioma de dominancia)
Axioma4: $A1 \circ A2 = A2 \circ A1$ (comut. débil)	C4: $A1 \bullet \geq A2 \Leftrightarrow A1 \circ A \bullet \geq A2 \circ A$, y $A1 \bullet \geq A2 \Leftrightarrow A \circ A1 \bullet \geq A \circ A2$	MRB4: $\forall (A \supset B, A \cap C = \emptyset) \Rightarrow (A \bullet \geq B \Rightarrow A \cup C \bullet \geq B \cup C)$ (monotonidad parcial)
Axioma5: $A1 \bullet \geq A2 \Rightarrow A1 \circ A \bullet \geq A2 \circ A$ (monotonidad débil)		MRB5: $\forall A \in \mathfrak{S}: A \bullet \geq 0$ (positividad)
Axioma6: Si $A3 \bullet > A4 \Rightarrow$ para cualquier $A1, A2$, \exists un natural n , tal que $A1 \circ nA3 \bullet > A2 \circ nA4$ (axioma arquimedeano)		
Sabiendo que una relación binaria $\bullet \geq$ es de orden débil si es transitiva y completa: $A1 \bullet \geq A2$, y $A2 \bullet \geq A3 \Rightarrow A1 \bullet \geq A3$ $A1 \bullet \geq A2$ o $A2 \bullet \geq A1$	Dónde $A1 = A2$ si y sólo si $A1 \bullet \geq A2$ y $A2 \bullet \geq A1$, y $A1 \bullet > A2$ si y sólo si $A1 \bullet \geq A2$ y no $(A2 \bullet \geq A1)$.	

Tabla 3. Estructuras del marco formal de Zuse [30]

Por último, debemos recordar que existen cinco tipos de escalas, que, de forma jerárquica, son: nominal, ordinal, intervalo, ratio y absoluta. Cada tipo de escala se define por transformaciones admisibles. Las medidas del software arrancan en la escala ordinal [30].

Seguidamente, se presenta la transformación que del marco formal se ha hecho para adaptarlo a las bases de datos relacionales, comprobando si las métricas propuestas por los autores verifican los distintos axiomas.

4 Caracterización de las métricas propuestas en el marco formal

En los sistemas de bases de datos relacionales, y para nuestros propósitos, el Sistema Relacional Empírico se puede definir como: $R = (R, \bullet \geq, \circ)$, donde R es un conjunto no vacío de relaciones (tablas), $\bullet \geq$ es la relación empírica "mas o igual complejo que" en R y \circ es una operación binaria (concatenación) cerrada en R .

En nuestro caso, la operación de concatenación será la combinación natural que se define de forma general como [12]: $Q \leftarrow R_{\langle \text{lista1} \rangle} \bullet \langle \text{lista2} \rangle S$, donde $\langle \text{lista1} \rangle$ especifica una lista de i atributos de R y $\langle \text{lista2} \rangle$ es una lista de i atributos de S . Estas listas son utilizadas para realizar las condiciones de comparación de igualdad entre pares de atributos. Después, estas condiciones son relacionadas mediante el operador AND. Solamente la lista correspondiente a la relación R se preserva en Q . La combinación natural, dependiendo de las características de las tablas que se combinan puede derivar en el producto cartesiano. Además, la combinación natural se puede realizar a través de una clave ajena y su clave principal o a través de dos columnas cualesquiera definidas sobre el mismo dominio. Será necesario tener todo esto en cuenta a la hora de diseñar las operaciones de combinación de cada una de las métricas.

4.1 Métrica NA

Mediante la figura 2 podemos ver las características que la métrica cumple y que nos ayudan a definir su función de combinación como: $NA(R_i \circ R_j) = NA(R_i) + NA(R_j) - NA(R_i \cap R_j)$, donde $NA(R_i \cap R_j)$ es el número de atributos comunes a (pertenecen a la intersección de) R_i y R_j .

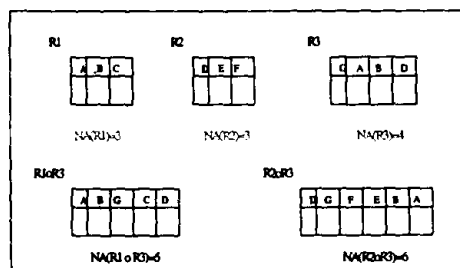


Figura 2. NA al combinar tablas

² Esta intersección es distinta a la operación de intersección entre dos relaciones del álgebra relacional.

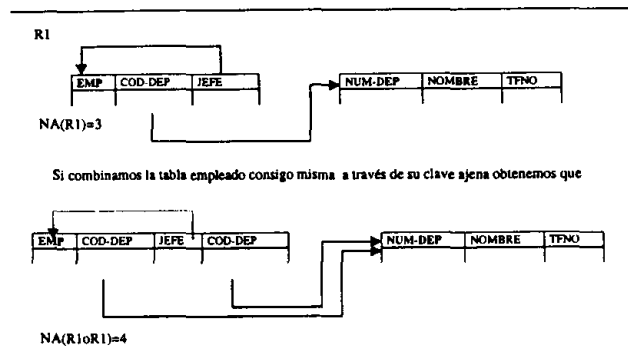


Figura 3. Idempotencia

Comprobamos si NA cumple los axiomas de la estructura extensiva modificada. El primer axioma (de orden débil) sí lo cumple ya que si tenemos dos relaciones R1 y R2, siempre se va a cumplir que $NA(R1) \geq NA(R2)$ o $NA(R2) \geq NA(R1)$ (compleción) y, siendo R1, R2 y R3 tres relaciones, también se va a cumplir que si $NA(R1) \geq NA(R2)$ y $NA(R2) \geq NA(R3)$, entonces $NA(R1) \geq NA(R3)$. NA en cambio, no cumple la positividad (axioma 2) porque si combinamos una relación R1 sin ciclos consigo misma $NA(R1 \circ R1)$ no va a ser mayor que $NA(R1)$. En cambio, la métrica sí cumple los axiomas 3 y 4 (asociatividad débil y conmutatividad débil) ya que la combinación natural es asociativa y conmutativa pero no cumple la monotonicidad débil (axioma 5) ya que como puede verse en la figura 2, al combinar R1 y R2 (tales que $NA(R1) = NA(R2)$) con R3 no se cumple que $NA(R1 \circ R3) \geq NA(R2 \circ R3)$. Para comprobar si cumple o no el axioma Arquimedeano, comprobamos previamente si la métrica es idempotente, ya que si lo es no cumple este axioma [30]. Para ello utilizaremos el ejemplo de la figura 3. Vemos que, en el caso presentado, al combinar una tabla consigo misma no permanece el número de atributos sino que aumenta. Luego no cumple la ley de idempotencia y hay que comprobar el axioma Arquimedeano. Para probar este axioma utilizamos el hecho de que al concatenar una tabla con otra tabla mediante combinación natural en sucesivas ocasiones, el número de atributos varía, y por tanto siempre se podrán encontrar unas A1, A3 tales que el axioma arquimedeano no se cumplirá, como muestra la figura 4 (dónde se presenta el resultado de la primera concatenación ya que las sucesivas concatenaciones obtendrán las mismas tablas $R1 \circ R3 \circ R3 \dots = R1 \circ R3$). Por tanto la métrica NA no cumple la estructura extensiva modificada.

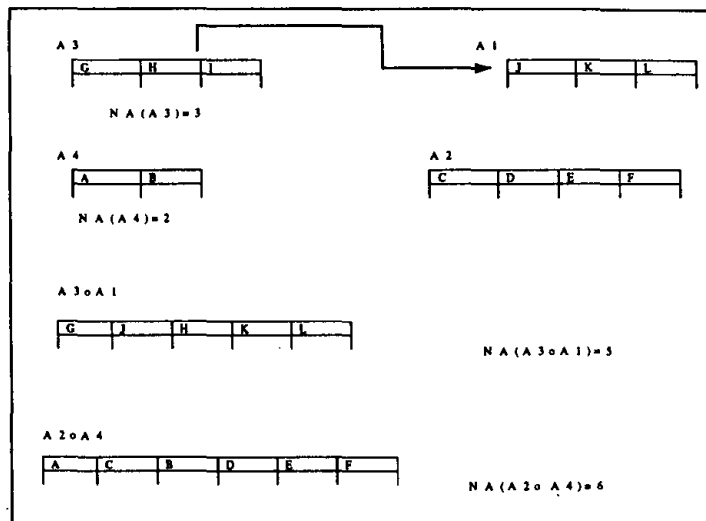


Figura 4. Axioma de Arquímedes

Comprobemos ahora si cumple las condiciones de independencia. La métrica no cumple la condición primera, como puede verse en la figura 2 ($NA(R1)=NA(R2)$ pero $NA(R1oR3)<NA(R2oR3)$). Si no se cumple la primera tampoco puede hacerlo la segunda. Al no cumplir el axioma 5 (monotonicidad débil) tampoco puede cumplir la tercera condición de independencia, y si no cumple la tercera, tampoco puede cumplir la cuarta. Por tanto tampoco cumple las condiciones de independencia.

Pasamos por tanto a comprobar la estructura de creencia (*belief*). MRB1 se cumple ya que dadas dos relaciones $R1$ y $R2 \in \mathfrak{S}$ (\mathfrak{S} es el conjunto de todas las posible relaciones hechas con los atributos del esquema relacional) $NA(R1) \geq NA(R2)$ o $NA(R2) \geq NA(R1)$. MRB2 también se cumple ya que la combinación natural es transitiva. Para comprobar si MRB3 se cumple consideramos una relación $R1 \supseteq R2$ (todos los atributos de $R2$ están en $R1$), por tanto $NA(R1) \geq NA(R2)$ y MRB3 se cumple. MRB4 también se va a cumplir ya si $R1 \supset R2$ (entonces $NA(R1) > NA(R2)$) y además $R1 \cap R3 = \emptyset$ (las relaciones $R1$ y $R3$ no tienen ningún atributo en común, y por tanto $R2$ y $R3$ tampoco) y añadimos atributos de $R3$ a $R1$ y a $R2$, entonces el número de atributos de $(R1oR3)$ será mayor que el número de atributos de $(R2oR3)$. Por último MRB5 también se va a cumplir ya que una relación debe tener siempre cero ó más atributos. Por tanto podemos concluir que cumple los axiomas de creencia (*belief*) lo cual concuerda con lo dicho por [30] respecto a las métricas que tienen función de creencia (como NA).

Como resultado podemos concluir que NA puede ser caracterizada como una medida por encima de la escala ordinal, y que asume la relación modificada de creencia (*belief*).

4.2 Métrica DRT

Para obtener la función de combinación de DRT hemos de pensar en las distintas posibilidades, de un parte tenemos que la combinación natural derive en producto cartesiano o se realice combinando columnas no relacionadas por integridad referencial,

en este caso los caminos referenciales no se ven modificados, correspondiendo el valor final de la métrica al máximo de las longitudes de los distintos caminos referenciales. Por otro lado tenemos la opción de concatenación mediante clave ajena, en la que puede verse o no afectada la longitud. En caso de verse afectada lo será decrementando una unidad. Así pues, y con el fin de generalizar podemos definir la regla de combinación como: $DRT(AoB) = \max(DRT(A), DRT(B)) - v$, siendo v una variable que puede tomar los valores 0 ó 1.

Veamos si DRT es una estructura extensiva modificada. Axioma 1. Sean $R1, R2$ y $R3$ tres tablas de un esquema de base de datos relacional, siempre se va a cumplir que: $DRT(R1) \geq DRT(R2)$ o $DRT(R2) \geq DRT(R1)$, y además se va a cumplir que: si $DRT(R1) \geq DRT(R2)$ y $DRT(R2) \geq DRT(R3) \Rightarrow DRT(R1) \geq DRT(R3)$, por lo que la métrica cumple el axioma 1. El axioma de positividad no se va a cumplir debido a la definición de la métrica (cuando v sea distinto de cero). La asociatividad y la conmutatividad, axiomas 3 y 4, se cumplen ya que la combinación natural es asociativa y conmutativa. El axioma 5 no tiene porque cumplirse, el ejemplo de la figura 5 nos lo demuestra. Antes de comprobar el axioma 6 comprobamos si la métrica es idempotente. Como es evidente, por mas que concatenemos una tabla consigo misma, nunca va a aumentar la longitud del camino referencial por lo que podemos afirmar que DRT es idempotente y que, por tanto, no va a cumplir el axioma de Arquímedes. Así pues podemos concluir que la métrica DRT no es una estructura extensiva modificada.

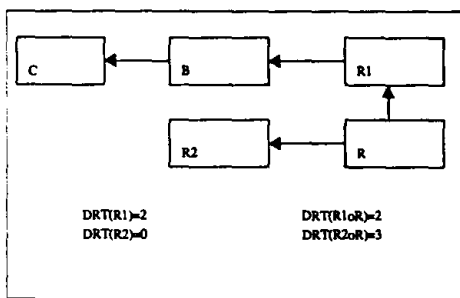


Figura 5. DRT no cumple el axioma 5

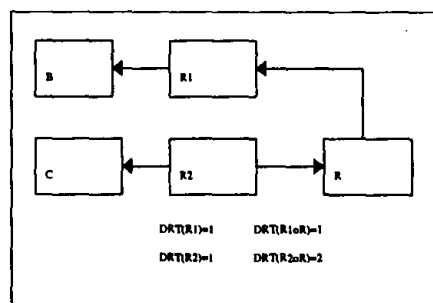


Figura 6. DRT no cumple C1

Veamos ahora si la métrica cumple las condiciones de independencia. Para comprobar que no cumple la condición primera utilizamos la figura 6. Al no cumplir C1 tampoco puede cumplir C2. Y por no cumplir el axioma 5 no puede cumplir la condición 3. Asimismo por no cumplir la tres tampoco podrá cumplir la 4.

Pasamos, por último, a comprobar si la métrica cumple las condiciones de una estructura de creencia. Las dos primeras propiedades corresponden a la compleción y a la transitiva, que se cumplen ya que la métrica pertenece a un orden parcial. La tercera se cumple ya que si todos los caminos referenciales de B están en A, entonces la longitud de A será como poco igual que la de B. La cuarta condición (monótona parcial) también se cumple ya que voy a ir añadiendo tanto a A como a B caminos referenciales de longitud máxima igual por lo que teniendo en cuenta que el DRT de A era mayor que el de B

antes de la concatenación también lo será después. La última condición (positividad) también se cumple ya que una longitud será como poco cero.

Como resultado podemos concluir que DRT puede ser caracterizada como una medida por encima de la escala ordinal, y que asume la relación modificada de creencia (*belief*).

4.3 Métrica RD

Para obtener la función de combinación de RD hemos de pensar en las distintas posibilidades, de un parte tenemos que la concatenación se realice mediante producto cartesiano o mediante columnas definidas sobre igual dominio, en este caso el número de claves ajenas no se ve modificado y correspondiendo el valor final de la métrica al número de claves ajenas totales. Por otro lado tenemos la opción de concatenación mediante clave ajena-clave principal, en la que se verá afectado el número de claves ajenas (disminuyendo en uno). Por tanto, la regla de combinación para la métrica RD será: $RD(R_i \circ R_j) = RD(R_i) + RD(R_j) - v$, dónde v puede tomar los valores 0 ó 1.

Comprobamos en primer lugar si cumple los axiomas de la estructura extensiva modificada. El primer axioma (de orden débil) sí lo cumple por que cumple la transitiva y la completión. El axioma de la positividad (axioma 2), también lo cumple ya que puede tomar únicamente los valores 0 o 1 y, por tanto, $RD(A1)+RD(A2)-v \geq RD(A1)$. La asociatividad débil (axioma 3) y la conmutatividad débil (axioma 4), también se cumplen ya que la combinación natural es asociativa y conmutativa. El axioma 5 (monotonicidad débil) no se cumple y lo podemos ver en el ejemplo de la figura 7. Vemos, por último el axioma Arquimedeano (axioma 6) que hay que comprobar ya que como podemos observar en la figura 3, al concatenar una tabla consigo misma se aumenta el número de claves ajenas y, por tanto, la métrica no es idempotente. Para comprobar que el axioma 6 no se cumple utilizamos la figura 8, dónde se muestra un ejemplo en el que el axioma no se cumple.

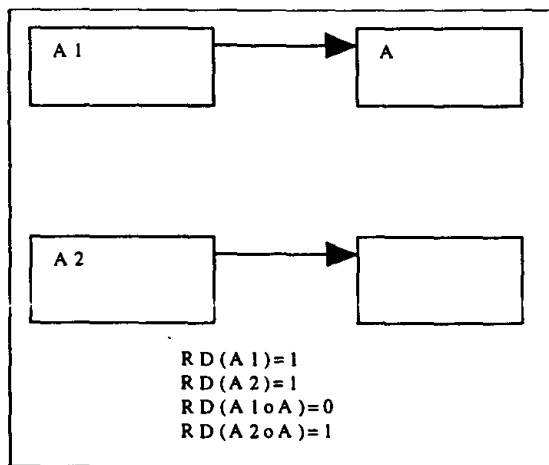


Figura 7. RD no cumple el axioma 5 de la estructura extensiva modificada

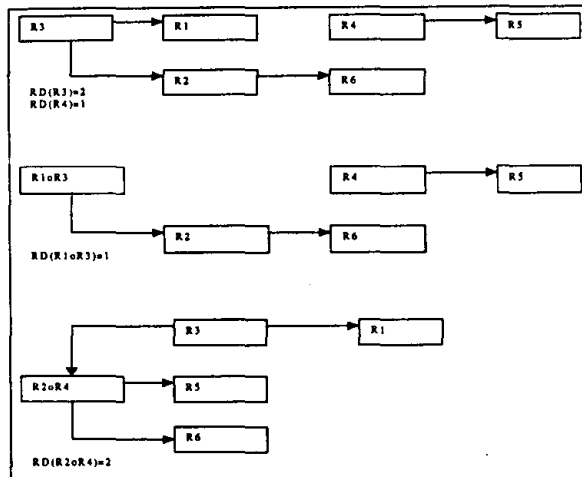


Figure 8. RD no cumple el axioma Arquimedeano

A continuación comprobamos si RD cumple las condiciones de independencia. La condición primera no se cumple ya que si observamos en la figura 8 podemos ver que R2 y R4 tienen un valor para la métrica igual ($RD=1$) y si combinamos ambas con R5, obtenemos que $RD(R2oR5)=1$ y que $RD(R4oR5)=0$. Si la primera no se cumple, la segunda tampoco va a cumplirse. Respecto a la tercera condición de independencia no puede cumplirse al no haberse cumplido el axioma 5 de la estructura extensiva modificada, y si no se cumple la tercera tampoco puede cumplirse la cuarta. Por tanto la métrica RD no cumple las condiciones de independencia.

Pasamos, por último, a comprobar la estructura modificada de creencia. Como RD cumple el orden débil también va a cumplir las condiciones primera y segunda de esta estructura. La tercera también se cumple ya que si todas las claves ajenas de B están en A, se va a cumplir que $RD(A) \geq RD(B)$. Estudiamos ahora la cuarta condición, si $A \supset B$ entonces $RD(A) > RD(B)$, si no hay claves ajenas comunes entre A y C, tampoco va a poder haberlas entre B y C, y por tanto, la condición cuarta se va a cumplir ya que $RD(AoC) > RD(BoC)$. La última condición también se cumple ya que el número de claves ajenas va a ser como mínimo cero pero nunca podrá tomar un valor menor a este.

Como resultado podemos concluir que RD puede ser caracterizada como una medida por encima de la escala ordinal, y que asume la relación modificada de creencia (*belief*).

5 Conclusiones y trabajo futuro

Creemos que es necesario abordar con mayor rigor los aspectos relativos a la calidad de las bases de datos relacionales, para lo que resulta imprescindible utilizar una serie de métricas específicas que guíen al diseñador, ayudando a elegir entre diferentes esquemas alternativos. Es necesaria una mayor investigación en métricas de software [24] tanto desde el punto de vista teórico como práctico [15].

En este artículo se ha propuesto un conjunto de métricas para medir la complejidad del esquema de bases de datos relacionales; estas métricas asumen, como las orientadas a objetos [30], propiedades más complejas relacionadas con la operación de concatenación que las métricas clásicas (de programas) y aunque estas métricas no cumplan las propiedades de la estructura extensiva modificada, pueden ser caracterizadas por encima de la escala ordinal por cumplir todas las propiedades de la estructura de creencia (*belief*). Esta clasificación resulta importante ya que es en la escala ordinal donde comienza la clasificación de las medidas del software (con números nominales, escala anterior, no pueden hacerse muchas operaciones ya que su transformación es limitada y, por lo tanto, las métricas resultarían poco útiles).

Pero no basta con comprobar si las métricas cumplen o no unas determinadas propiedades matemáticas ya que puede que una métrica sea correcta desde un punto de vista formal, pero no tener relevancia práctica para un problema determinado. Es por esto por lo que resultan necesarias validaciones empíricas que nos permitan comprobar y entender las implicaciones de las medidas en nuestros productos y procesos. Esto se consigue a través de planteamiento de hipótesis en el mundo real, mas allá de la pura teoría, que habrá que comprobar con datos empíricos. Basándonos en esta necesidad de estudios empíricos, realizamos varios experimentos con las métricas relacionadas con la integridad referencial (DRT y RD). Nuestro objetivo era demostrar que estas métricas pueden ser utilizadas para medir la complejidad del esquema de bases de datos relacional, la cual influye en su entendibilidad. Los resultados obtenidos nos permitieron llegar a las conclusiones siguientes: el número de claves ajenas (RD) en el esquema de base de datos relacional es un indicador sólido de su entendibilidad y la longitud del camino referencial (DRT) por sí mismo no resulta indicativo, aunque modula los efectos en combinación con RD [4].

Este trabajo debe ser complementado con casos de estudio, con el fin de superar las limitaciones que presentan los experimentos controlados [1]. Por todo ello, en la actualidad estamos llevando a cabo una evaluación más profunda de estas y otras métricas en colaboración con empresas y organismos en situaciones reales. Por otro lado, estamos construyendo una herramienta que permite calcular de forma automática diversas métricas, para SGBD basados en SQL92, de manera que podamos dar una guía más precisa para la utilización de las mismas [5]. Asimismo estamos elaborando métricas para bases de datos activas [11] y objeto-relacionales [26].

Bibliografía

- 1 Basili, V. R., (1999), Using experiments to build a body of knowledge. Madrid. Spain.
- 2 Batini, C., Ceri, S. y Navathe, S., (1992), *Conceptual database design. An entity relationship approach*. Benjamin Cummings Publishing Company.
- 3 Briand, L., Morasca, S. y Basili, V. Property-Based (1996), Software Engineering Measurement. *IEEE Transactions on software Engineering*, Vol. 22, No 1, January.
- 4 Calero, C., Piattini M., Polo, M y Ruiz, F. (1999). Validating referential integrity as a database quality metric. *ICEIS 99*, 27-31 March. Portugal.
- 5 Churcher, N.J. y Shepperd, M.J. (1995), Comments on "A Metrics Suite for Object-Oriented Design. *IEEE Trans. on Software Engineering* Vol. 21, N° 3, 1995, pp. 263-265.

- 6 Codd, E.F. (1970). A Relational Model of Data for Larged Shared Data Banks. *CACM*, 13 (6), 377-387.
- 7 Connolly, T., Begg, C. y Strachan, A., (1998), *Database systems*, 2ª Ed. Addison-Wesley
- 8 Date, C.J. (1995). *An Introduction to Database Systems*. 6th. Addison-Wesley, Reading, Massachusetts
- 9 De Miguel, A., Piattini, M., (1993), *Concepción y diseño de bases de datos relacionales*, Ed. RA-MA
- 10 De Miguel, A., Piattini, M., Marcos, F. (1999), *Diseño de bases de datos relacionales*, Ed. RA-MA
- 11 Díaz, O and Piattini, M. (1999) Metrics for active databases maintainability. *Accepted in CAISE'99*. Heidelberg, June 16-18.
- 12 Elmasri R y Navathe S. (1997). *Database Systems*. Second edition. Addison-Wesley, Massachussets
- 13 Fenton, N. (1994). Software Measurement: A Necessary Scientific Basis. *IEEE Transactions on Software Engineering*, 20(3): 199-206.
- 14 Fenton, N. y Pfleeger, S.L., (1997), *Software Metrics: A rigorous approach*, 2ª edición. International Thomson Computer Press.
- 15 Glass, R. (1996). The Relationship Between Theory and Practice in Software Engineering. *IEEE Software*, November, 39 (11), 11-13.
- 16 Gray R.H.M., Carey B.N., McGlynn N.A. y Pengelly A.D., (1991), Design metrics for database systems. *BT Technology J*, Vol 9, 4 Oct, 69-79
- 17 Hansen, G.W. y Hansen J.V., (1997), *Diseño y administración de bases de datos*, 2ª Ed. Ed. Prentice-Hall
- 18 Henderson-Sellers, B. (1996). *Object-oriented Metrics - Measures of complexity*. Prentice-Hall, Upper Saddle River, New Jersey
- 19 ISO, (1994). *Software Product Evaluation-Quality Characteristics and Guidelines for their Use*. ISO/IEC Standard 9126, Geneva.
- 20 Li, H.F. y Chen, W.K. (1987). An empirical study of software metrics. *IEEE Trans. on Software Engineering*, 13 (6): 679-708.
- 21 MacDonell, S.G., Shepperd, M.J. y Sallis, P.J. (1997). Metrics for Database Systems: An Empirical Study. *Proc. Fourth International Software Metrics Symposium - Metrics'97*, Albuquerque. IEEE Computer Society, pp. 99-107.
- 22 McCabe, T.J. (1976). A complexity measure. *IEEE Trans. Software Engineering* 2(5): 308-320.
- 23 McFadden, F.R., Hoffer, J.A. y Prescott, M.B., (1999), *Modern database management*, 5ª Ed. Addison-Wesley
- 24 Neil, M. (1994) Measurement as an Alternative to Bureaucracy for the Achievement of Software Quality. *Software Quality Journal* 3 (2), 65-78.
- 25 Pfleeger, S. L. (1997). Assessing Software Measurement. *IEEE Software*. March/April, pp. 25-26.
- 26 Piattini, M., Calero, C., Polo, M. y Ruiz, F. (1998). Maintainability in Object-Relational Databases. *Proc of The European Software Measurement Conference FESMA 98*, Antwerp, May 6-8, Coombes, Van Huysduynen and Peeters (eds.), 223-230
- 27 Sneed, H.M. y Foshag, O. (1998). Measuring Legacy Database Structures. *Proc of The European Software Measurement Conference FESMA 98*, Antwerp, May 6-8, Coombes, Van Huysduynen and Peeters (eds.), 199-211.
- 28 Stonebraker, M. y Brown, J., (1999), *Object-Relational DBMSs: Tracking the next great wave*. 2ªEd. Morgan Kaufmann Publishers Inc.
- 29 Weyuker, E.J. (1998), Evaluating software complexity measures. *IEEE Transactions on Software Engineering* Vol. 14, N° 9, pp. 1357-1365, Sept.
- 30 Zuse, H. (1998). *A Framework of Software Measurement*. Berlin, Walter de Gruyter.