

ICIE 2000

Entregado por ICIE 2000

Completado por J. M. Skrlec, A. Pauluzzi
Department School of Engineering
City of Buenos Aires

ISSN 887-481764-7



ICIE 2000

Facultad de Ingeniería
Información, Ingeniería

de Abril de 2000

Facultad de Ingeniería
Buenos Aires - Argentina

Prefacio
Autoridades
Auspicios Institucionales
Indice

ISBN 987-461764-7

El camino sigue siendo la investigación.

El Congreso Internacional de Ingeniería Informática (ICIE) del Departamento de Computación de la Facultad de Ingeniería de la Universidad de Buenos Aires, ya forma parte de las actividades de Investigación en Ciencias de la Computación, INFORMÁTICA, de las Universidades, Facultades y Laboratorios de la Argentina e Internacionales de la Región y otros Centros de todo el mundo, que basan su excelencia académica en la generación del conocimiento y el intercambio de datos y métodos científicos y tecnológicos.

El Departamento sigue en su marcha hacia el asentamiento de sus unidades investigativas, laboratorios.

Estos laboratorios funcionan como aglutinantes naturales de profesores, docentes auxiliares, graduados (en ejercicio de la profesión) y alumnos avanzados en torno a proyectos de investigación y desarrollo.

Los profesores y docentes auxiliares con dedicación exclusiva o semiexclusiva asocian su plan de investigación a alguna de las líneas de investigación de los laboratorios.

Esta filosofía hace que los alumnos puedan elegir el tema de su tesis de graduación de una carpeta de proyectos adscriptos a las diferentes líneas de investigación de los laboratorios. A partir de este momento, el alumno queda vinculado al laboratorio como asistente de investigación y el docente responsable del proyecto por el cual optó el alumno ejerce las funciones de dirección del proyecto que comienza el alumno.

Los laboratorios se articulan con las orientaciones de la Carrera de Grado y los Posgrados.

Esta estructura de investigación tiene algunas ventajas:

- Provee una oferta de líneas de investigación muy amplia
- Permite, a través de la modalidad de los proyectos, disponer de la exploración de hipótesis de investigación alternativas en forma simultánea.
- Permite compartimentalizar las tareas de investigación:

Si los productos específicos de la enseñanza son los que van a integrar los planteles o crear empresas con gran capacidad de adaptación y aplicación de nuevos conocimientos que generan la actual revolución tecnológica y la diferenciación para cubrir todos los nichos del mercado, podríamos pensar que la adecuación del actual sistema universitario es condición necesaria para producir un profesional del próximo siglo.

Han aparecido varios problemas, y tal vez el menos trivial sea cómo se realimenta la Universidad de los requerimientos de la sociedad.

Los grupos de trabajo, de interés y laboratorios son el fundamento del desarrollo futuro en la industria.

La filosofía encarada, considera que en la Enseñanza y la Investigación las líneas de acción tienen que ser puras, la construcción de sistemas no debe mezclarse con los paradigmas de programación ya que el problema pasa por la arquitectura, la calidad y la modularidad de los sistemas informáticos y sus componentes.

Se debe estimular el pensamiento científico, es decir, estimular la visión analítica, resolver un problema y luego, en alguna actividad intermedia del ciclo de vida tener una visión ingenieril de la informática.

La Sociedad requiere nuevas habilidades de los profesionales, que las Universidades como formadoras deben dar respuesta. La capacidad de poder articularse con los centros de generación de conocimiento es una de ellas, *pero definitivamente se deben respetar las incumbencias de la*

Informática.

El problema de la formación de recursos humanos con excelencia académica, se ataca mediante la creación de laboratorios y su fortalecimiento.

Prof. Gregorio Perichinsky.

Director del Departamento de Computación.

Miembro del Comité de Programa.

AUTORIDADES DEL COMITÉ DE PROGRAMA



PROF. GREGORIO PERICHINSKY

*Facultad de
Ingeniería,
Universidad de Buenos Aires, Argentina*



PROF. DANIEL BORRAJO

*Universidad Carlos III de Madrid,
Españ*

PROF. ARACELI PROTO

*Comisió de
Investigaciones
Cienficas,
Argentina*

PROF. RAIMUNDO D'AQUILA

*ITBA, Instituto
Tecnológico
Buenos Aires, Argentina*

PROF. MAURICIO TORO

*Universidad de Sevilla,
Españ*



PROF. PEDROA ISASI VIÑUELA

*Universidad Carlos III de Madrid,
Españ*

PROF. GREGORY RANDALL

*Universidad de la
República*



PROF. JORGE ARDENGHI

Universidad Nacional del Sur, Argentina



PROF. RAÚL GALLARD

Universidad Nacional de San Luis, Argentina



PROF RODOLFO BERTONE

Universidad Nacional de La Plata, Argentina



PROF. STELLA MARIS VALIENTE

Universidad Nacional de Mar del Plata, Argentina



PROF. ANTONIO MARSIGLIO

Universidad Nacional de Mar del Plata, Argentina

PROF. BIBIANA D. ROSSI

Escuela Metropolitana de Altos Estudios, Argentina



PROF. NELSON BALOIAN

Universidad Nacional de Chile, Chile

o Ejecutivo

PROF. FERNANDO JAVIER LAGE



*Facultad de
Ingeniería,
Universidad de Buenos Aires*

■
PROF. ZULMA CATALDI

*Facultad de
Ingeniería,
Universidad de Buenos Aires*

SRTA. ANALÍA PAULUZZI

*Facultad de
Ingeniería,
Universidad de Buenos Aires*

SR. JUAN SKRILEC

*Facultad de
Ingeniería,
Universidad de Buenos Aires*

AUTORIDADES DEL COMITÉ EVALUADOR



PROF. ANTONIO PÉREZ AMBITE

UNIVERSIDAD POLITÉCNICA DE MADRID, ESPAÑA



PROF. ARMANDO DE GIUSTI

UNIVERSIDAD NACIONAL DE LA PLATA, ARGENTINA



PROF. PATRICIA PESADO

UNIVERSIDAD NACIONAL DE LA PLATA, ARGENTINA



PROF. OMAR IGLESIAS

UNIVERSIDAD NACIONAL DE LA PLATA , ARGENTINA



PROF. RAÚL PESSACQ

UNIVERSIDAD NACIONAL DE LA PLATA , ARGENTINA



PROF. RODOLFO BERTONE

UNIVERSIDAD NACIONAL DE LA PLATA , ARGENTINA



PROF. OSCAR BRÍA

UNIVERSIDAD NACIONAL DE LA PLATA , ARGENTINA



PROF. CLAUDIA VILLAR

UNIVERSIDAD NACIONAL DE LA PLATA , ARGENTINA

PROF. ARACELI PROTO (CIC)

PROF. MARÍA CRISTINA DONNAMARÍA (CIC)



PROF. LEOPOLDO CARRANZA

***FACULTAD DE
INGENIERÍA
- UNIVERSIDAD DE BUENOS AIRES, ARGENTINA***



PROF. ZULMA CATALDI

***FACULTAD DE
INGENIERÍA
- UNIVERSIDAD DE BUENOS AIRES, ARGENTINA***



PROF. FERNANDO J. LAGE

***FACULTAD DE
INGENIERÍA
- UNIVERSIDAD DE BUENOS AIRES, ARGENTINA***



PROF. ARTURO CARLOS SERVETTO

***FACULTAD DE
INGENIERÍA
- UNIVERSIDAD DE BUENOS AIRES, ARGENTINA***



PROF. MARÍA FELDGEN

***FACULTAD DE
INGENIERÍA
- UNIVERSIDAD DE BUENOS AIRES, ARGENTINA***



PROF. OSVALDO CLÚA

***FACULTAD DE
INGENIERÍA
- UNIVERSIDAD DE BUENOS AIRES, ARGENTINA***



PROF. JULIA DENAZIS

***FACULTAD DE
INGENIERÍA
- UNIVERSIDAD DE BUENOS AIRES, ARGENTINA***

PROF. BIBIANA ROSSI

EMAE

PROF. RAIMUNDO D'AQUILA

INSTITUTO TECNOLÓGICO BUENOS AIRES, ARGENTINA

PROF. PAOLA BRITOS

INSTITUTO TECNOLÓGICO BUENOS AIRES, ARGENTINA



PROF. JORGE R. ARDENGHI

UNIVERSIDAD NACIONAL DEL SUR, ARGENTINA



PROF. CARLOS RICO

UNIVERSIDAD NACIONAL DE MAR DEL PLATA, ARGENTINA



PROF. ANTONIO MARSIGLIO

UNIVERSIDAD NACIONAL DE MAR DEL PLATA, ARGENTINA

INGENIERÍA SOFTWARE

UTILIZACIÓN DEL MÉTODO FORMAL CÁLCULO DE SISTEMAS DE COMUNICAC*Raúl Gómez**González**Universidad de La Serena, La Serena - CHILE***ASPECTOS DO PROJETO E IMPLEMENTAÇÃO DE AMBIENTES MULTIPARADIGM***Aparecido Valdemir de FREITAS, João**JoseNETO**Escola Politécnica da Universidade de**São Paulo, Depto. de Engenharia de**Computação**e Sistemas Digitais, Cidade Universitária**São Paulo, BRASIL***CALIDAD EN EL MODELADO DE DATOS***Mario Piattini, Marcela Genero, Coral Calero e Ismael Caballero**Grupo ALARCOS - Departamento de Informática. Escuela Superior de**Informática**Universidad de Castilla-La Mancha, ESPAÑA***RELATIONSHIPS AS OBJECT TYPES: IMPROVING CONCEPTUAL MODELLING***Esperanza Marcos**Univ. Rey Juan Carlos, ESPAÑA***ESPECIFICANDO FORMALMENTE APRESENTAÇÕES MULTIMÍDIA INTERATIVA:***Alessandra M. de Castro Porto, Rosvelter Coelho da Costa**Universidade do Vale do Itajaí Universidade Federal de Santa Catarina, BRASIL***FERRAMENTAS DE AUTORIA DE REALIDADE VIRTUAL - UM ESTUDO COMPARA***Adhemar Maria do Valle Filho, Patricia Cristiane de Souza, João Bosco da Mota Alves, Raul Sidn**Universidade Federal de Santa Catarina -UFSC, Universidade do Vale do**Itajaí-**UNIVALI,**Faculdades Integradas Cândido Rondon**-UNIRONDON**Florianópolis**-SC**BRASIL***A FORMAL BASIS FOR RECURSIVE PATTERNS***Amaro S., Martinez Carod N., Cechich A.**Departamento de Informática y**Estadística, Universidad Nacional del Comahue**Buenos Aires, Neuquen, ARGENTINA*

**CONCEITOS DE CASOS DE USO E ENTIDADE E RELACIONAMENTO COMO FERR**

Jussara Pimenta MATOS e Selma Shin Shimizu MELNIKOFF.

Departamento de Engenharia de

Computaçã

e Sistemas Digitais

Escola Politécnica, Universidade de

Sã Paulo, BRASIL

**BASES PARA EL DISEÑO DE UN REPOSITORIO IMPLEMENTADO EN UN CASE DI**

María Delia Grossi, Elizabeth

Jiménez Rey, Gregorio Perichinsky

Laboratorio de Bases de Datos y Sistemas Operativos

Departamento de Computació - Facultad de

Ingeniería

Universidad de Buenos Aires

CALIDAD EN EL MODELADO DE DATOS¹

Mario Piattini, Marcela Genero, Coral Calero e Ismael Caballero

Grupo ALARCOS

Departamento de Informática. Escuela Superior de Informática

Universidad de Castilla-La Mancha

Ronda de Calatrava, 5 - 13071 Ciudad Real

{mpiattin, mgenero, ccalero, icaballero}@inf-cr.uclm.es

La calidad es un tema emergente en la actualidad no sólo en los ámbitos económicos y organizativos sino también en el mundo de las bases de datos y de los Sistemas de Información. En los últimos años, se han publicado algunos trabajos que abordan temas relacionados con la evaluación de la calidad en las bases de datos teniendo en cuenta desde la calidad de los modelos de datos hasta la calidad de los propios datos. En este artículo mostramos una visión general del estado del arte en cuanto a la calidad del modelado de datos, revisando también algunas de las métricas propuestas en este sentido.

1. INTRODUCCIÓN

Debido a la creciente complejidad de los sistemas de información, es necesario dedicar una mayor atención a la evaluación de la calidad de las bases de datos a lo largo del proceso de desarrollo, no sólo para satisfacer a los usuarios sino también para mejorar la productividad del propio proceso de desarrollo (Moody y Shanks, 1994).

La mayoría de los trabajos realizados con respecto a la calidad se refieren a la calidad del software (Gillies, 1992; Arthur, 1992; Oskarsson y Glass, 1996; Jones, 1997; ISO, 1994; Ginac, 1998). La calidad de las bases de datos es un tema que ha sido descuidado en el ámbito de la investigación (Sneed y Foshag, 1998). Incluso en el diseño tradicional de las bases de datos, los aspectos referidos a la calidad no se incorporan explícitamente (Wang et al, 1993).

En la calidad de las bases de datos se pueden identificar diferentes aspectos, tal y como se muestra en la figura 1: la calidad del modelo conceptual de datos, la calidad del SGBD utilizado para implementar dicha base de datos y también la calidad de los propios datos.

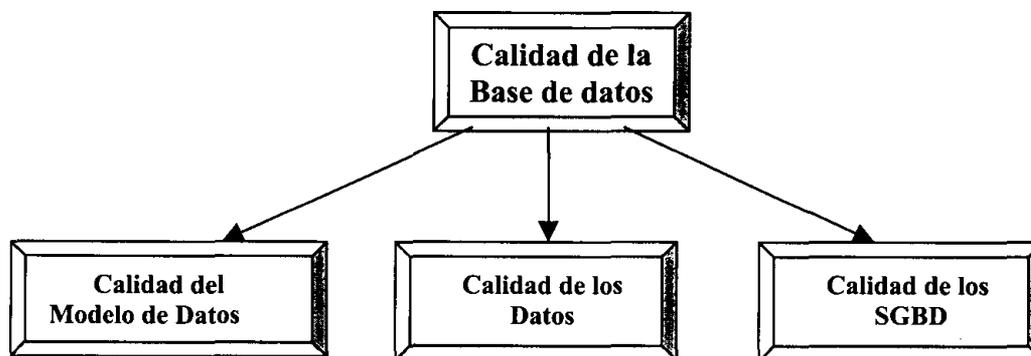


Figura 1. Calidad en bases de datos

¹ Este trabajo se encuadra dentro de los proyectos MANTICA, parcialmente financiado por la CICYT y la Unión Europea (1FD97-0168) y CALIDAT, que desarrolla la empresa Cronos Ibérica, S.A. en colaboración con la Universidad de Castilla-La Mancha con financiación de la Consejería de Educación y Cultura de la Comunidad de Madrid (Ref: 09/0013/1999).

En este artículo describiremos diferentes propuestas que han sido desarrolladas para analizar la calidad de los modelos conceptuales de datos; presentando asimismo las diferentes métricas propuestas. Por lo que respecta a la calidad de los datos, merece la pena destacar los trabajos de Redman (1996), Orr (1998), Bobrowski et al. (1999), English (1999) y Huang et al. (1999). En cuanto a la calidad de los SGBD utilizados para implementar las bases de datos, consideramos que no es inherente a la base de datos y que es difícilmente modificable por los usuarios, pudiendo ser abordada como la calidad de cualquier producto software utilizando la norma ISO 9126 (ISO, 1994).

2. CALIDAD DE MODELOS CONCEPTUALES DE BASES DE DATOS

Aunque la etapa de modelado de datos representa solo una proporción pequeña del esfuerzo total del desarrollo de sistemas, el impacto sobre el resultado final es probablemente mayor que el de cualquier otra etapa (Simsion, 1991), de ahí la importancia de asegurar la calidad del modelado.

Algunos autores (Batini et al., 1992; Reingruber y Gregory, 1994; Boman et al., 1997) han definido la calidad de los modelos conceptuales como una lista de propiedades “deseables” (por ejemplo: compleción, corrección, minimalidad, expresividad, legibilidad, autoexpli-cación, extensibilidad y normalidad) que debe satisfacer el modelo de datos y han propuesto una serie de transformaciones con el objetivo de mejorar la calidad de los mismos. Aunque estas listas proveen un punto de partida útil para entender y mejorar la calidad en el modelado de datos, la mayoría no son estructuradas, usan definiciones imprecisas, a menudo se solapan, mezclándose características de la especificación con las del método y del lenguaje de modelado; además de presuponer la existencia de diseño/implementación, y se presentan objetivos poco realistas o imposibles de alcanzar (Lindland et al., 1994).

Una estrategia más adecuada para abordar la calidad es definir marcos de referencia que organicen y estructuren los conceptos clave y características en el modelado de datos. Recientemente, se han publicado algunos marcos de referencia que siguen esta línea como Moody y Shanks (1994), Moody et al. (1998), Shanks y Darke (1997), Lindland et al. (1994), Krogstie et al. (1995), Eick (1991), Kesh (1995) y Schuette y Rotthowe (1998). En la tabla 1 se clasifica y compara las propuestas más recientes de acuerdo a su propósito, características, tipo y enfoque (Piattini et al., 2000).

2.1 PROPUESTA DE LINDLAND ET AL.

Este marco, basado en la teoría semiótica, se presentó en Lindland et al. (1994) con el objetivo de paliar las deficiencias detectadas en los enfoques de listas de propiedades, al mismo tiempo que se separan los objetivos de calidad de los medios para alcanzarlos y se utiliza un fundamento matemático en su descripción. En Krogstie et al. (1995) se ha extendido este marco con el fin de incorporar el concepto de acuerdo social de Pohl (1994), con lo que se pueden identificar los siguientes elementos (véase figura 2):

- Audiencia: unión del conjunto de actores individuales, el conjunto de actores sociales organizacionales y el conjunto de actores técnicos que necesitan relacionarse con el esquema.
- Modelo: conjunto de todas las sentencias expresadas explícita o implícitamente.
- Lenguaje: conjunto de todas las sentencias que se pueden expresar de acuerdo al vocabulario y la gramática de los “lenguajes de modelado” (modelos) utilizados.
- Dominio: conjunto de todas las sentencias que serían correctas y relevantes acerca del problema.

Autor(es)	Objetivo	Características	Tipo	Enfoque
Roman (1985)	Definir propiedades para especificaciones de requisitos	Propiedades asociadas a su utilización en el proceso de diseño	Lista de propiedades	Teórico
Gray et al. (1991)	Diseñar bases de datos de buena calidad	Métricas para evaluar la complejidad de esquemas E/R	Lista de métricas	Práctico
Eick (1991)	Mejorar la calidad de los modelos conceptuales	Métrica para medir la calidad global de los diagramas S.	Lista de métricas	Práctico
Batini et al. (1992)	Mejorar la calidad de los esquemas conceptuales	Características de calidad de un buen esquema	Lista de propiedades	Práctico
Lindland et al. (1994)	Entender la calidad en el modelado conceptual	Base lingüística.	Marco de referencia	Teórico
Simsion (1994)	Características de calidad esquemas E/R	Diseño y evaluación de esquemas alternativos	Lista	Práctico
Pohl (1994)	Definir objetivos y dimensiones de proceso para modelado de req.	Especificación, representación y acuerdo	Marco de referencia	Teórico
Moody y Shanks (1994)	Evaluar la calidad del modelo E/R	Factores de calidad, Estrategias y métodos de evaluación	Marco de referencia	Práctico
Reingruber y Gregory (1994)	Construir modelos conceptuales de buena calidad	Características de calidad de un buen esquema	Lista de propiedades	Práctico
Levitin y Redman (1995)	Características de calidad esquemas E/R	Dimensiones de calidad,	Lista de propiedades	Práctico
Krogstie et al. (1995)	Entender la calidad en el modelado conceptual	Extensión de Lindland et al.	Marco de referencia	Teórico
Kesh (1995)	Evaluar la calidad de los modelos E/R	Define métricas par evaluar c/u de los factores de calidad	Marco de Referencia	Práctico
Boman et al. (1997)	Construir modelos conceptuales de buena calidad	Características de calidad de un buen esquema	Lista de propiedades	Práctico
Shanks y Darke (1997)	Entender la calidad en los modelos conceptuales	Un marco integrado	Marco de Referencia	Teórico-Práctico
Moody et al. (1998) y Moody (1998)	Evaluar la calidad del modelo E/R	Extensión del marco de Moody y Shanks (1994). Propone 25 métricas para medir cada factor de calidad.	Marco de Referencia	Práctico
Schuette y Rotthowe (1998)	Manejar la subjetividad del modelado	Una Guía de Modelado, con principios y objetivos asociados	Marco de Referencia	Práctico
Polo et al. (1998)	Evaluar calidad de esquemas E/R	Extensión de Batini et la. (1992)	Lista de métricas	Práctico
Genero et al. (1999)	Evaluar complejidad de modelos de objetos OMT.	Conjunto de métricas de complejidad	Lista de métricas	Práctico
Piattini et al. (2000)	Evaluar la complejidad de modelos E/R	Conjunto de métricas acotadas	Lista de métricas	Práctico

Tabla 1.- Enfoques en la calidad del modelado conceptual de datos

- Interpretación de la audiencia: conjunto de todas las sentencias de las que la audiencia piensa que consta el modelo.
- Conocimiento de los participantes: unión de los conjuntos de sentencias de todos los actores sociales individuales.

Como es natural, se considera que un esquema tiene una mayor calidad semántica cuanto mejor sea la correspondencia entre el esquema conceptual diseñado y el dominio que se pretende representar. Sin embargo, como señalan los autores de este marco, es imposible establecer o verificar directamente esta correspondencia, ya que para diseñar el esquema se debe acudir al conocimiento que tiene el público sobre el dominio y para verificarlo se debe emplear la interpretación que el público hace del esquema.

En este marco se propone distinguir varios tipos de calidad:

- **Calidad sintáctica:** que viene determinada por la correspondencia entre el modelo y el lenguaje (modelo), y cuyo objetivo es la corrección sintáctica.
- **Calidad semántica:** (percibida), que comprende tanto la validez (lo expresado en el modelo es correcto y relevante para el problema) como la completión (el modelo contiene todas las sentencias acerca del dominio que son correctas y relevantes).
- **Calidad pragmática:** cuyo objetivo es que el modelo sea comprendido.
- **Calidad social:** que persigue distintos tipos de acuerdo tanto en la interpretación del modelo como respecto al conocimiento del dominio.

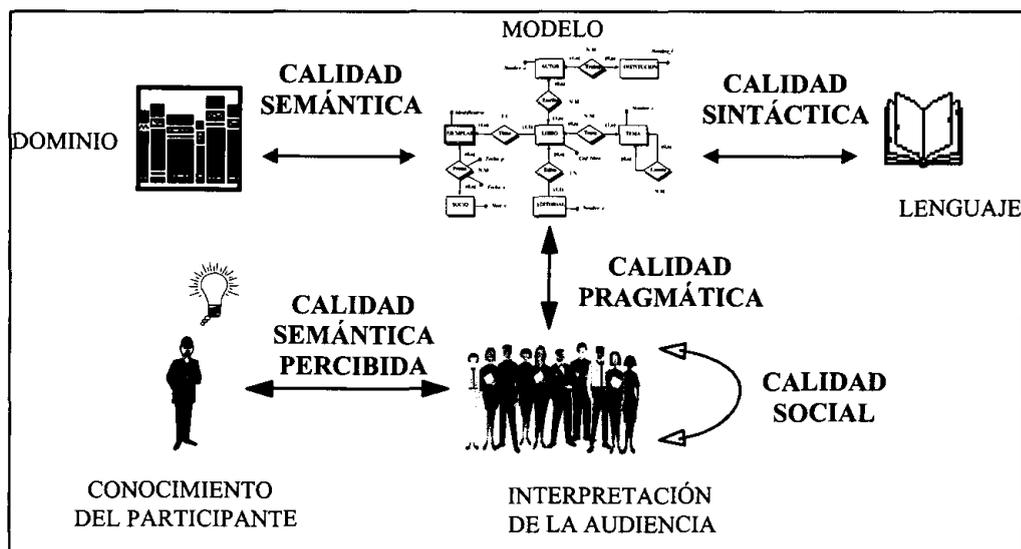


Figura 2. Marco para la calidad del modelado conceptual, Krogstie et al. (1995)

Este marco propone además distintos medios para mejorar los diferentes tipos de calidad. Aunque el marco presentado y sus posteriores extensiones, constituyen una buena contribución para mejorar la comprensión de los aspectos de la calidad concernientes al modelado conceptual, tienen un alto nivel de abstracción y no son directamente aplicables en la práctica.

2.2 PROPUESTA DE MOODY Y SHANKS

Otro marco, con un enfoque más práctico que el anterior, fue presentado por Moody y Shanks (1994) y refinado recientemente en Moody et al. (1998). Pretende ayudar a los diseñadores a la hora de elegir entre distintas alternativas de diseño y de poder acomodar las distintas visiones de los distintos implicados (“*stakeholders*”) en el proceso de modelado de datos. Este marco presenta los siguientes elementos relacionados con el esquema (véase figura 3):

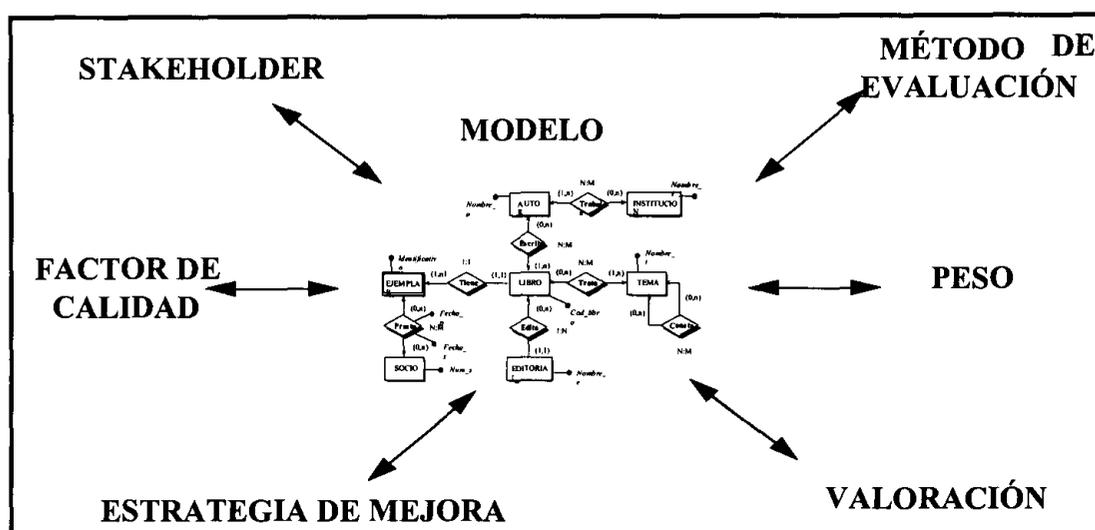


Figura 3. Elementos del Marco de Moody y Shanks (1994)

- Factor de calidad: propiedad deseable de un esquema conceptual.
- Implicado (*stakeholder*): persona involucrada en la construcción o utilización del esquema.
- Estrategia de mejora: técnica para mejorar la calidad de los esquemas conceptuales.
- Método de evaluación: modo sistemático de evaluar factores de calidad.
- Peso: que sirve para definir la importancia relativa de los factores de calidad.
- Valores: representan la valoración de un factor de calidad por alguno de los implicados.

En la última revisión de este marco (Moody et al., 1998), se proponen ocho factores de calidad: completión, integridad, flexibilidad, comprensibilidad, corrección, simplicidad, integración e implementabilidad. Este marco se ha validado en varios casos prácticos, en los que también se ha estudiado la influencia que ejercen unos factores sobre otros; así, por ejemplo, aumentar la implementabilidad del esquema puede acarrear una disminución de su flexibilidad y completión.

Shanks y Darke (1997) han propuesto integrar el marco de Moody y Shanks (1994) con el de Lindland et al. (1994), ya que ambos se complementan (este último desde un punto de vista más teórico, mientras que el de Moody y Shanks se enfoca más hacia la práctica) y además comparten conceptos: los implicados (*stakeholders*) podrían asimilarse al público, los objetivos y las propiedades a los factores de calidad y el concepto de modelo es compatible en los dos marcos (véase figura 4).

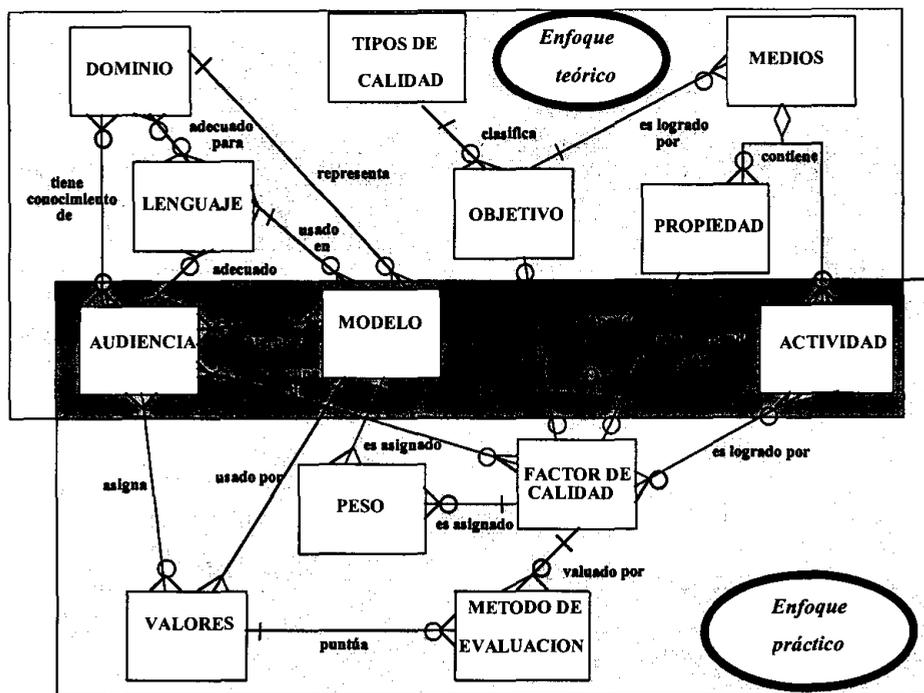


Figura 4. Metamodelo propuesto por Shanks y Darke (1997)

2.3 PROPUESTA DE KESH

Kesh (1995) propuso el desarrollo de un modelo, una metodología y métricas para evaluar la calidad del modelo E/R. El modelo está basado en conceptos tomados de la inteligencia artificial (específicamente de los sistemas basados en el conocimiento) y la ingeniería de software. La figura 5 muestra el marco de referencia utilizado para evaluar la calidad, que considera dos tipos de componentes: de funcionamiento y ontológicos.

Según este autor, los modelos E/R están formados por dos componentes ontológicos: la *estructura* y el *contenido*. La *estructura* se refiere a las entidades y sus relaciones mientras que el *contenido* se refiere a los atributos de las entidades.

La tabla 2 muestra como los factores ontológicos afectan a los factores de comportamiento.

		UU	UD	M	P	R
Estructura	Adecuación al problema	X				
	Validez		X	X		
	Consistencia	X	X		X	
Contenido	Concisión	X		X		X
	Compleción	X	X		X	X
	Cohesión		X	X		
	Validez		X			

UU = usabilidad (usuario) UD= usabilidad(diseñador) M = manten P = precisión R= rendimiento

Tabla 2. Modelo para evaluar la calidad de los modelos E/R

2.4 PROPUESTA DE SCHUETTE Y ROTTHOWE

La propuesta de Schuette y Rotthowe (1998) se basa en la suposición de que la postura subjetiva del modelador es un aspecto relevante para el resultado del modelado conceptual y por lo tanto se debe manejar y comprender tal subjetividad. Siguiendo ciertas reglas durante el proceso de modelado es posible construir modelos subjetivamente comparables. Especialmente en proyectos donde intervienen muchos diseñadores tales reglas son absolutamente necesarias. Estos autores definen:

- Una Guía de modelado (GoM) obtenida a partir de los problemas que surgen del proceso subjetivo de diseño de un sistema GoM contiene seis principios que permiten mejorar la calidad del modelado de la información: Principio de adecuación de la construcción, principio de adecuación del lenguaje, principio de la eficiencia económica, principio de claridad, principio del diseño sistemático y principio de comparabilidad
- La arquitectura GoM, que es un marco estructural en el cual se ubican cada uno de los componentes de esta guía.

3. MÉTRICAS PARA LA CALIDAD DE MODELOS CONCEPTUALES

3.1 MÉTRICAS DE MOODY

En la tabla 3 se detallan 25 métricas propuestas en Moody (1998) algunas se calculan de forma objetiva (p. ej. número de entidades del esquema), mientras que otras resultan de la puntuación subjetiva de los implicados en el diseño.

Factor de calidad	Métricas
Compleción	<ul style="list-style-type: none"> – N° de elementos del modelo de datos que no corresponden con requisitos de usuario – N° de requisitos de usuario no representados en el modelo de datos – N° de elementos de datos que corresponden a requisitos de usuario pero definidos de forma inexacta – N° de inconsistencias con el modelo de procesos
Integridad	<ul style="list-style-type: none"> – N° de reglas del negocio que no se hacen cumplir por el modelo de datos – N° de restricciones de integridad incluidas en el modelo de datos que no corresponden a políticas del negocio
Flexibilidad	<ul style="list-style-type: none"> – N° de elementos en el modelo que están sujetos a cambios en el futuro – Costes estimados de los cambios – Importancia estratégica de los cambios
Comprensibilidad	<ul style="list-style-type: none"> – Valoración de los usuarios sobre la comprensibilidad del modelo – Capacidad de los usuarios de interpretar el modelo correctamente – Valoración de los desarrolladores de aplicaciones sobre la comprensibilidad del modelo
Corrección	<ul style="list-style-type: none"> – N° de violaciones de las convenciones de modelado de datos – N° de violaciones a las formas normales – N° de instancias de redundancia en el modelo
Simplicidad	<ul style="list-style-type: none"> – N° de entidades – N° de entidades e interrelaciones – N° de constructores ($aN^E + bN^R + cN^A$)
Integración	<ul style="list-style-type: none"> – N° de conflictos con el modelo de datos corporativo – N° de conflictos con los sistemas existentes – Valoración de los representantes de todas las áreas de negocio
Implementabilidad	<ul style="list-style-type: none"> – Valoración de riesgo técnico – Valoración de riesgo de planificación – Estimación del coste de desarrollo – N° de elementos físicos incluidos en el modelo de datos

Tabla 3. Métricas de Moody (1998) para evaluar la calidad de los modelos E/R.

3.2 MÉTRICAS DE KESH

Según Kesh (1995), una vez que se termina de construir un diagrama E/R se debe evaluar su calidad utilizando la siguiente metodología:

- 1) Calcular el valor de cada componente ontológico en forma individual.
- 2) Para componente de comportamiento combinar los valores de los componentes ontológicos que le son relevantes.
- 3) Combinar los valores de los componentes de comportamiento para calcular el valor de la calidad.
- 4) Si el valor de un componente ontológico se encuentra por debajo de un cierto valor, debe ser investigado y mejorado antes de seguir adelante.

El valor global de la calidad Q se calcula de la siguiente manera:

$$Q = w_1 \cdot s_1 + w_2 \cdot s_2 + w_3 \cdot s_3 + w_4 \cdot s_4 + w_5 \cdot s_5$$

siendo w_1, w_2, w_3, w_4, w_5 los pesos asociados a cada factor de comportamiento, usabilidad(usuario), usabilidad(diseñador), mantenibilidad, exactitud y rendimiento y $s_1, s_2, s_3, s_4, s_5, s_6$ los valores de cada factor de comportamiento. Para determinar el valor de Q es necesario estimar dichos pesos y valores.

El valor de los pesos puede ser determinado o bien por la importancia del factor para la organización o por un acuerdo general o estándar acerca de los valores (comúnmente no existen). Los valores s_1, s_2, s_3, s_4 y s_5 , se calculan como el promedio de los factores ontológicos:

$$\begin{aligned} s_1 &= (o_1 + o_3 + o_4 + o_5) / 4 \\ s_2 &= (o_2 + o_3 + o_5 + o_6 + o_7) / 5 \\ s_3 &= (o_2 + o_4 + o_6) / 3 \\ s_4 &= (o_3 + o_5) / 2 \\ s_5 &= (o_4 + o_5) / 2 \end{aligned}$$

Donde, por ejemplo, para calcular o_1 (adecuación del esquema al problema) se entrevista a los usuarios para determinar si el esquema se adecua al problema, y éstos lo puntúan con una escala de 1 a 5.

3.3 MÉTRICAS DE EICK

Eick (1991) propuso una métrica para evaluar la calidad de los modelos conceptuales, que ha sido integrada dentro de una metodología denominada ANNAPURNA, implementada en una herramienta CASE. Los esquemas conceptuales planteados por este autor se denominan “diagramas S” y están basados en el modelo relacional binario y el SDM (Semantic Database Model), e incluyen clases, conexiones entre clases, atributos, jerarquías de generalización (superclase/subclases), “dependencias de existencia” (que se dan cuando la existencia de un conjunto de atributos implica la existencia de otro), “dependencias funcionales” (presentes si dos entidades que tienen ciertos atributos iguales también tienen que concordar en otros).

Dado un universo de discurso U y un conjunto de diagramas S ($\{S_1, S_2, \dots, S_N\}$) que describen dicho universo de discurso, y siendo:

- qu_1 , el número de dependencias funcionales que se dan en U pero no se expresan en S_i
- qu_2 , el número de dependencias en existencia que se dan en U pero no se expresan en S_i
- qu_3 , el número de atributos y conexiones de subtipos de los esquemas S_i
- qu_4 , el número de clases de S_i

- qu_5 , el número de etiquetas de S_i
- f es una función de N^5 a R^+ (siendo N el conjunto de los números naturales y R^+ el de los reales positivos)

La calidad de S relativa a U se define como $f(qu_1, qu_2, qu_3, qu_4, qu_5)$:

$$f = \frac{1}{3 * qu_1 + 3 * qu_2 + qu_3 + qu_4 + qu_5}$$

3.4 MÉTRICAS DE POLO ET AL.

En Polo et al. (1998) se ha definido un marco completo para la medición de atributos de calidad en las etapas de diseño conceptual siguiendo la filosofía de la norma ISO 9126 (ISO, 1994), que considera la calidad como la agregación de varios factores. El conjunto de factores de calidad que se emplean utiliza los propuestos en Batini et al. (1992) complementados con otros nuevos, todos ellos enfocados hacia la mantenibilidad de las bases de datos. Así, por ejemplo, se definen, entre otras:

- El ratio de minimalidad como:

$$MR = 1 - \frac{NDA}{NAS} \quad \dots \text{donde:}$$

NDA es el número de atributos repetidos, deducibles mediante operaciones o derivables por inferencia lógica del esquema
NAS es el número de atributos del esquema

- La normalidad, basada en la métrica Área propuesta en Gray (1991), como:

$$N = \frac{Ae \times Ee + Re \times Ae + Re \times Ee}{A3 \times E3 + R3 \times A3 + R3 \times E3}$$

...donde el numerador es calculado en función del número de atributos, entidades y relaciones del esquema real, y el denominador lo es en función del mismo en tercera forma normal
A = número de atributos, E = número de entidades y R = número de relaciones

- El ratio de cohesión del esquema”:

$$COSR = \frac{\sum_{i=1}^{|US|} NEUS_i^2}{NES^2}$$

$|US|$ es el número de componentes conexas del grafo representado por el esquema conceptual
 $NEUS_i$ es el número de entidades en la componente conexa i -ésima
 NES es el número de entidades del esquema.

3.5 MÉTRICAS DE GENERO ET AL.

En Piattini et al. (1999) se han definido un conjunto de métricas que permiten medir un factor de la calidad que es la “simplicidad”, aplicadas a los modelos E/R. Todas estas métricas están basadas en el concepto de métricas *acotadas* (closed-ended) (Lethbridge, 1998). Una métrica acotada es aquella en la cual las mediciones siempre caen dentro de un rango (por ejemplo, $[0,1]$). Las métricas acotadas son más útiles que las no acotadas, porque sus valores están acotados dentro de un rango y por consiguiente su significado es más relevante. Permiten a los diseñadores hacer comparaciones entre distintos modelos E/R alternativos y obtener resultados objetivos acerca de su calidad.

Estas métricas están basadas en la teoría de la complejidad, que define a la complejidad de un sistema como el número de componentes del sistema y el número de relaciones entre los componentes. Como la meta es simplificar el modelo E/R tanto como sea posible, el objetivo será minimizar el valor de estas métricas, conservando el contenido semántico del esquema E/R. Consideramos que si el valor de una métrica tiende a cero es el mejor caso, y si tiende a uno el peor caso.

Algunas de las métricas consideradas en este trabajo son:

- **Métrica RvsE**

La métrica Relaciones vs. Entidades mide la complejidad introducida por el número relaciones que hay en un modelo E/R con respecto al número de entidades. Está basada en la métrica M_{RPROP} propuesta por Lethbridge (1998).

Esta métrica se define de la siguiente manera:

$RvsE = \left(\frac{N^R}{N^R + N^E} \right)^2$	N^R es el número de relaciones. N^E es el número de entidades. Siendo $N^R + N^E > 0$.
-------------------------------------------------	---------------------------------------------------------------------------------------------------

- **Métrica DA**

La métrica Atributos Derivados mide la complejidad introducida por los atributos derivados. Se define como:

$DA = \frac{N^{DA}}{N^A - 1}$	N^{DA} es el número de atributos derivados. N^A es el número total de atributos. Siendo $N^A > 1$.
-------------------------------	---------------------------------------------------------------------------------------------------------------

- **Métrica CA**

La métrica Atributos Compuestos mide la complejidad introducida por los atributos compuestos en un modelo E/R. Se define como:

$CA = \frac{N^{CA}}{N^A}$	N^{CA} es el número de atributos compuestos. N^A es el número de atributos. Siendo $N^A > 0$.
---------------------------	----------------------------------------------------------------------------------------------------------

- **Métrica MVA**

La métrica Atributos Multivaluados mide la complejidad introducida por los atributos multivaluados en un modelo E/R. Se define como:

$MVA = \frac{N^{MVA}}{N^A}$	N^{MVA} es el número de atributos multivaluados. N^A es el número de atributos. Siendo $N^A > 0$.
-----------------------------	--------------------------------------------------------------------------------------------------------------

- **Métrica RR**

La métrica Relaciones Redundantes mide la complejidad introducida por las relaciones que son redundantes, en un modelo E/R. Se define como:

$RR = \frac{N^{RR}}{N^R - 1}$	N^{RR} número de relaciones redundantes. N^R número de relaciones. Siendo $N^R > 1$.
-------------------------------	-------------------------------------------------------------------------------------------------

- **Métrica M:NRel**

La métrica relaciones M:N mide la complejidad introducida por la relaciones M:N en el modelo E/R. Se define como el número de relaciones M:N que hay en el modelo E/R dividido el número total de relaciones que hay en dicho modelo.

Esta métrica se define de la siguiente manera:

$M : N Rel = \frac{N^{M:NR}}{N^R}$	$N^{M:NR}$ es el número de relaciones M:N. N^R es el número de relaciones. Siendo $N^R > 0$.
------------------------------------	-------------------------------------------------------------------------------------------------------

- **Métrica N-aryRel**

La métrica Relaciones N-arias mide la complejidad introducida por las relaciones N-arias (no binarias) en el modelo E/R. La definimos como:

$N - \text{ary Rel} = \frac{N^{N-\text{ary}R}}{N^R}$	$N^{N-\text{ary}R}$ es el número de relaciones N_arias. N^R es el número de relaciones. Siendo $N^R > 0$.
------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------

- **Métrica IS_ARel**

La métrica Relaciones IS_A evalúa la complejidad de las jerarquías de generalización/especialización (jerarquías IS_A) en un modelo E/R. Está basada en la métrica M_{ISA} definida por Lethbridge (1998). Esta métrica combina dos factores con el objetivo de medir la complejidad de las jerarquías de herencia. El primer factor es la fracción de entidades en la jerarquía de generalización/especialización, que son hojas. Esta medida, llamada Fleaf la calculamos así:

$FLeaf = \frac{N^{Leaf}}{N^E}$	N^{Leaf} es el número de entidades que son hojas en una jerarquía de generalización/especialización. N^E es el número total de entidades en una jerarquía de generalización/especialización Siendo $N^E > 0$.
--------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

La figura 5 muestra diferentes jerarquías de herencia con sus respectivos valores de Fleaf. Fleaf tiende al valor 0,5 cuando el número de hojas es la mitad del número de entidades (figura 5, ej. c y d). Tiende a cero en el caso de un árbol unario (figura 5, ej. b). Y tiende a 1 cuando cada entidad es un subtipo de la entidad raíz de la jerarquía (figura 5, ej. a). Fleaf tiene una propiedad poco deseable, para jerarquías poco profundas (dos o tres niveles) con una gran cantidad de hojas nos da un valor demasiado alto (figura 5, ej. a). Para corregir este problema, utilizamos un factor adicional para calcular el valor de la métrica IS_ARel, el promedio de los supertipos directos o indirectos para cada entidad que no sea la raíz. A este factor lo denominamos ALLSup (la entidad raíz no la tenemos en cuenta ya que no tiene padres).

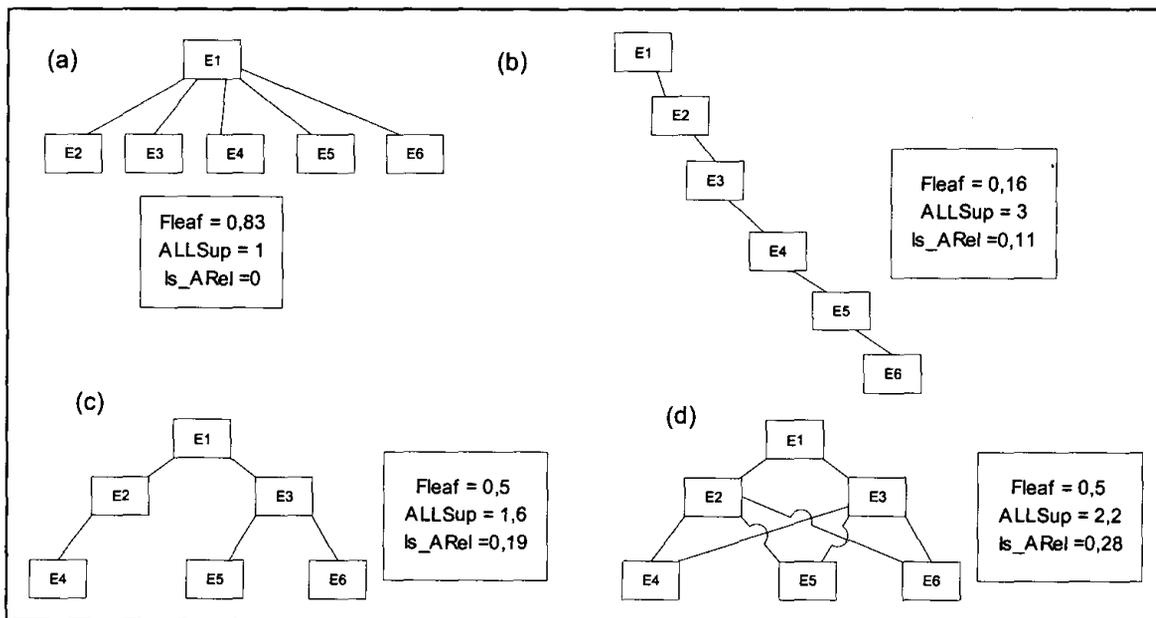


Figura 5. Ejemplos de relaciones IS_A

La métrica IS ARel se define de la siguiente manera:

$$Is_ARel = FLeaf - \frac{FLeaf}{ALLSup}$$

4. CONCLUSIONES

En este artículo hemos planteado la importancia de abordar el tema de la calidad en las bases de datos y hemos mostrado los trabajos de investigación realizados en esta área, que han tenido una mayor repercusión. Si bien todas estas propuestas han significado una gran contribución, creemos que es necesaria una mayor investigación en los aspectos relacionados con la calidad de las bases de datos y, especialmente, con la elaboración y refinamiento de métricas, tanto desde el punto de vista teórico como práctico.

Además de evaluar la calidad de los modelos conceptuales es necesario tener en cuenta aspectos de la calidad referidos a los modelos lógicos de bases de datos. Según nuestro conocimiento, hasta el momento se ha realizado muy poca investigación al respecto. Nosotros estamos realizando algunos trabajos sobre métricas para bases de datos relacionales (Calero et al., 1999; Polo et al., 1998), bases de datos objeto-relacionales (Piattini et al., 1998) y bases de datos activas (Díaz y Piattini, 1999).

Queremos recalcar, que hay que poner especial cuidado al definir métricas, ya que muchas veces no miden los atributos que pretenden medir (Briand et al, 1996). Una medición efectiva pasará por disponer de métricas definidas con rigor y validadas formalmente (Fenton, 1994; Morasca y Briand, 1997), además de que demuestren su validez empírica, tanto en experimentos controlados como en casos reales (Basili, 1999).

Además, es necesario profundizar en la investigación sobre la calidad del proceso, tanto del proceso de modelado como del proceso de obtención y carga de los datos, ya que la mayor parte de los estudios publicados se centran en la calidad del producto.

REFERENCIAS

- Arthur, L. (1992). *Improving Software Quality*. John Wiley & Sons, Inc.
- Basili, V.. (1999). *Using experiments to build a body of knowledge*. Conferencia impartida en la U.P.M.
- Batini, C., Ceri, S. y Navathe, S. (1992). *Conceptual database design. An entity relationship approach*. Benjamin Cummings Publishing Company.
- Boman M, Bubenko J., Johannesson P. y Wangler B. (1997). *Conceptual Modelling*. Prentice Hall.
- Briand, L., Morasca, S. y Basili, V. (1996). Property-Based Software Engineering Measurement. *IEEE Transactions on Software Engineering*, 22 (6), 68-86.
- Bobrowski, M., Marré, M. y Yankelevich, D. (1999). Measuring Data Quality. Informe Técnico n° 99-002, Departamento de Computación, Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.
- Calero, C., Piattini, M., Polo, M. y Ruiz, F. (1999) Validating referential integrity as a database quality metric. *Proceedings of the First International Conference on Enterprise Information Systems*, Portugal, 45-50.
- Díaz, O. y Piattini, M. (1999). Metrics for active databases maintainability. *Proc. CAISE'99*. Heidelberg, June 16-18.

- Eick, C. (1991). A Methodology for the Design and Transformation of Conceptual Schemas. *Proc. of the 17th International Conference on Very Large Data Bases*. Barcelona.
- English, L. (1999). *Improving Data Warehouse and Business Information Quality*. John Wiley & Sons, Inc.
- Fenton, N. (1994). Software Measurement: A Necessary Scientific Basis. *IEEE Transactions on Software Engineering*, 20(3), 199-206.
- Genero, M., Manso, M^a E., Piattini, and García, F. (1999). Assessing the Quality and the Complexity of OMT Models. *2nd European Software measurement Conference - FESMA 99*, Amsterdam, The Netherlands, 99-109.
- Gilles, A. (1992). *Software Quality- Theory and Managment*. Chapman & Hall Computing. London.
- Ginac, F. (1998). *Customer Oriented Software Quality Assurance*. Prentice Hall PTR, New Jersey.
- Huang, K-T., Lee, Y. y Wang, R. (1999). *Quality Information and Knowledge*. Prentice Hall, Upper Saddle River.
- ISO (1994). *Software Product Evaluation-Quality Characteristics and Guidelines for their Use*. ISO/IEC Standard 9126, Geneva.
- Jones, C. (1997). *Software Quality. Analysis and guidelines for success*. International Thomson Computer Press, Boston.
- Kesh, S. (1995). Evaluating the Quality of Entity Relationship Models. *Information and Software Technology*, 37 (12), 681-689.
- Krogstie, J., Lindland, O. I. y Sindre, G. (1995). Towards a Deeper Understanding of Quality in Requirements Engineering, *Proceedings of the 7th International Conference on Advanced Information Systems Engineering (CAISE)*, Jyvaskyla, Finland, June, 82-95.
- Lethbridge, T. (1998). Metrics For Concept-Oriented Knowledge bases. *International Journal of Software Engineering and Knowledge Engineering*, 8(2), 161-188.
- Lindland, O., Sindre, G. y Solvberg, A. (1994). Understanding Quality in Conceptual Modelling, *IEEE Software*, 11(2), Marzo, 42-49.
- Moody, D. (1998). Metrics For Evaluating the Quality of Entity Relationship Models. *Proceedings of the Seventeenth International Conference on Conceptual Modelling (E/R '98)*, Singapore, November 16-19, 213-225.
- Moody, D. y Shanks G. (1994). What Makes A Good Data Model? Evaluating The Quality of Entity Relationships Models. *Proceedings of the 13th International Conference on Conceptual Modelling (E/R '94)*, Manchester, England, December 14-17, 94-111.
- Moody, D., Shanks G. y Darke P. (1998). Improving the Quality of Entity Relationship Models – Experience in Research and Practice. *Proceedings of the Seventeenth International Conference on Conceptual Modelling (E/R '98)*, Singapore, November 16-19, 255-276.
- Morasca, S. y Briand, L.C. (1997). Towards a Theoretical Framework for measuring software attributes. *Proceeding of the Fourth International, Software Metrics Symposium*, 119-126.
- Orr, K. (1998). Data Quality and System Theory. *Communications of the ACM*, 41 (2), 66-71
- Oskarsson, Ö. y Glass, R. (1996). *An ISO 9000 Approach to building Quality Software*. Prentice Hall PTR, New Jersey.

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.

Piattini, M., Genero, M., Calero, C., Polo, M. y Ruiz, F. (2000). Metrics for Managing Quality in Information Modeling. "Information Modelling in the Next Millennium". Rossi, M y Siau, K. (eds.) Artech House, Londres.

Pohl, K. (1994). The Three Dimensions of Requirements Engineering: A Framework and its Applications, *Information Systems*, 19, 243-258.

Polo, M., Calero, C., Ruiz, F. y Piattini, M. (1998). Métricas de calidad y complejidad para bases de datos. *Actas de las III Jornadas en Ingeniería del Software*, Toval, A. y Nicolás, J. (eds.), 79-90.

Redman, T. (1996). *Data Quality for the Information Age*. Artech House, Boston.

Reingruber, M. y Gregory, W. (1994): *The Data Modelling Handbook. A best-practice approach to building quality data models*. John Wiley & Sons, Inc.

Schuette, R. y Rotthowe, T. (1998). The Guidelines of Modeling – An Approach to Enhance the Quality in Information Models. *Proceedings of the Seventeenth International Conference on Conceptual Modelling (E/R '98)*, Singapore, November 16-19, 240-254.

Shanks, G. y Darke, P. (1997). Quality in Conceptual Modelling: Linking Theory and Practice. *Proc. of the Pacific Asia Conference on Information Systems (PACIS'97)*, Brisbane, 805-814.

Simsion, G. Creative Data Modelling. *Proceedings of the Tenth International Entity Relationship Conference*, San Francisco, 1991.

Sneed, H. y Foshag, O. (1998). Measuring Legacy Database Structures. *Proc. of The European Software Measurement Conference FESMA '98*, Coombes, Hooft and Peeters (eds.), 199-210.

Wang, R., Kon, H. y Madnick, S. (1993). Data Quality Requirements Analysis and Modeling. *Proc. of the 9th International Conference on Data Engineering*, Viena, IEEE Computer Society, 670-677.

UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE COMPUTACIÓN

Por cuanto

Mario Piattini

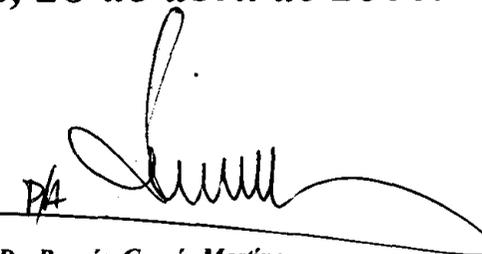
Ha participado del VI Congreso Internacional de Ingeniería Informática

Se le confiere el siguiente Certificado que así lo acredita.

Buenos Aires, 28 de abril de 2000.



Lic. Gregorio Perichinsky
Director
Departamento de Computación



Dr. Ramón García Martínez
Jefe División Investigación y Desarrollo
Departamento de Computación