

Grupo de Usuarios Oracle de España

CUORE



- Trampa en el Cyberespacio.
- Corba 2.0: la evolución hacia los "Business Objects".
- Cómo desarrollar una Intranet con tecnología Oracle.
- Asegurar la calidad de las bases de datos... un reto para el año 2000.

CUORE



Asegurar la calidad de las Bases de Datos un reto para el año

2000



MARCELA GÉNERO, CORAL CALERO, MARIO PIATTINI
(mgenero, calero, mpiattini@inf-ucrdm.es
Grupo ALARCOS

Departamento de Informática, Escuela Superior de Informática
Universidad de Castilla-La Mancha
Ronda de Calatrava, 5 - 13071 Ciudad Real

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, que son uno de los componentes más importantes 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 éste área.

1 INTRODUCCIÓN

En la actualidad, los temas relacionados con la calidad han pasado a ocupar un primer plano no solo en ámbitos económicos y organizativos, sino también en el mundo de las bases de datos que constituyen el corazón de los Sistemas de Información. En un mercado cada vez más competitivo y globalizado, la información hoy en día, juega un papel fundamental en las empresas, y la calidad de dicha información tiene una gran influencia sobre la toma de decisiones que permite a los directivos

plantear diferentes estrategias para alcanzar una mayor competitividad en el mercado.

Debido a la creciente complejidad de los sistemas de información, es necesaria una continua atención en la evaluación de la calidad de las bases de datos a lo largo del proceso de desarrollo, para tener conocimiento de la calidad de los sistemas. El compromiso de la calidad en el desarrollo de cualquier producto de software es esencial, no solo para satisfacer a los clientes sino también para mejorar la productividad del proceso de desarrollo (Moody y Shanks, 1994).

Existe un rango desconcertante de definiciones formales e informales con respecto al concepto de calidad (Gilles, 1992). La mayoría de las definiciones provistas han

surgido de ámbitos económicos y de fabricación. No debemos olvidar que la calidad es un concepto relativo, dado que la importancia de los diferentes factores que influyen en la calidad de un producto varía entre los diseñadores a lo largo del tiempo (Freund, 1985). Por consiguiente es importante que cualquier método de evaluación de la calidad reconozca estas diferencias de puntos de vista, y permita que los diferentes aspectos tengan diferentes pesos de acuerdo al contexto o situación que se esté considerando.

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, 1998; Ginac, 1998). La calidad de las bases de datos es un tema que ha

DIVULGACIÓN

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 son incorporados explícitamente Wang et al (1993). Aunque el trabajo con las bases de datos tradicionalmente no ha estado centrado en temas referidos al manejo de la calidad en sí misma, muchas de las herramientas desarrolladas le han dado importancia al manejo de la calidad de los datos. En efecto, la investigación ha estado dirigida a temas

referidos a como prevenir las inconsistencias de los datos (restricciones de integridad y teoría de la normalización) y a cómo prevenir la corrupción en los datos -manejo de transacciones (Wang et al., 1993).

Nosotros opinamos que en la calidad de las bases de datos influyen diferentes "calidades", 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.

En este artículo nos centraremos en los aspectos de la calidad de las bases de datos referidos a la calidad del producto, es decir, describiremos diferentes propuestas que han sido desarrolladas para abordar la calidad de los modelos conceptuales de datos y la calidad de los datos, empezando por esta última. No trataremos la calidad de los SGBD utilizados para implementar las bases de datos, ya que consideramos que no es inherente específicamente a la base de datos y que es difícilmente modificable por los usuarios.

2 CALIDAD DE DATOS

Prácticamente cada autor define su propio marco a la hora de caracterizar la calidad de los datos. Así,

por ejemplo, English (1999) destaca dos tipos de cuestiones relativas a la calidad de los propios datos:

- **Calidad inherente**, es decir la precisión de los datos, el grado en que los datos reflejan exactamente los objetos del mun-

do. Como sabemos, la calidad de un producto depende del proceso mediante el cual se diseña y produce el producto, así que la calidad de los datos estará relacionada con el proceso de captura, de diseño, o de utilización de los mismos. En Wang y Wang

(1996) se analizan algunas de las causas de la mala calidad de los datos debida a deficiencias en diseño basándose en principios ontológicos, véase *tabla 1*.

Hay que tener en cuenta, desde un punto de vista práctico, que las mayores deficiencias en la calidad

de los datos provienen de su no utilización (Orr, 1998). Al igual que pasa con los miembros y órganos del cuerpo humano, los datos que no se utilizan terminan "atrofiándose", y su calidad se deteriora.

Las empresas tendrán que, por un lado, definir una política de calidad que establezca las obligaciones de cada rol (en general, directivo de informática, creadores y suministradores de datos, diseñadores, operadores, etc.) con el fin de asegurar la calidad de los datos en todas sus dimensiones; mientras que por otro deberán implementar un proceso con el fin de evaluar la calidad de la información de que disponen.

Dentro de este proceso resulta fundamental disponer de medidas o métricas que permitan analizar y mejorar la calidad. En Huang et al. (1999) se propone considerar diferentes tipos de métricas: subjetivas (basadas en el juicio de los usuarios de los datos), objetivas independientes de la aplicación (como la corrección) y objetivas dependientes de la aplicación (específicas para un dominio determinado).

Por otro lado, hay algunas interesantes propuestas (Wang et al., 1995) sobre extender las bases de datos relacionales con indicadores que permitan asignar parámetros objetivos y subjetivos a la calidad de los valores de

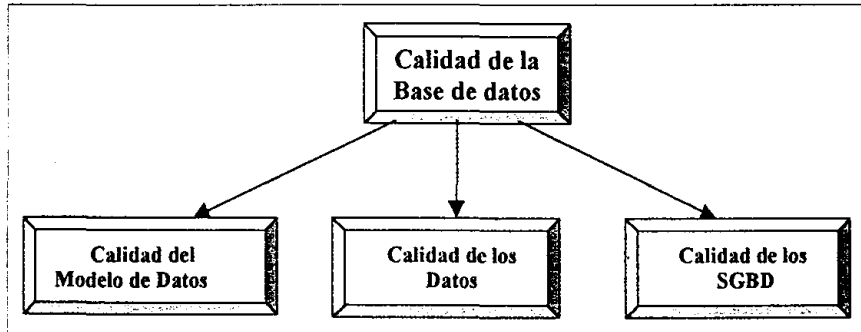


FIGURA 1 • Vista de la Calidad en las Bases de Datos

do real que representan, que abarcaría: conformidad con la definición, completión de valores, validez o conformidad con las reglas del negocio, precisión con la fuente, precisión con la realidad, no duplicación, accesibilidad

- **Calidad pragmática**, el grado en que los datos permiten a los "trabajadores del conocimiento" satisfacer los objetivos de la empresa de forma eficaz y eficiente: oportunidad, claridad contextual, integridad de derivación, usabilidad, corrección o completión de hechos.

Mientras que Redman (1997) señala cuatro dimensiones en la calidad de los datos:

- Precisión
- Completión
- Actualidad
- Consistencia de valores

Dimensión Calidad de Datos	Naturaleza de la deficiencia	Fuente de la deficiencia
Completión	Representación impropia: Estados SI ¹ ausentes	Fallo en el diseño
No ambigüedad	Representación impropia: Varios estados del MR ² mapeados al mismo estado SI	Fallo en el diseño
Significación	Estados SI sin sentido y confusión: mapeo a estado sin sentido	Fallo en el diseño y fallo en la operación
Corrección	Confusión: mapeo a un estado incorrecto	Fallo en la operación

TABLA 1 • Calidad de los datos y deficiencias de diseño

¹SI = Sistema de Información

²MR = Mundo Real

DIVULGACIÓN

la base de datos. Así, por ejemplo, en la tabla que se muestra a continuación se almacena para cada valor de la base de datos la fuente que proporciona el dato y la fecha en la que lo hace, además se debería conocer la credibilidad de la fuente (que en este caso podría ser alta), lo que ayudaría a los "trabajadores del conocimiento" en la toma de decisiones.

En Wang et al. (1993) se presenta incluso una metodología que complementa a las tradicionales metodologías de diseño de bases de datos. En la primera etapa, además de crear el esquema conceptual, por ejemplo, utilizando el modelo entidad/interrelación, se deberían identificar los requisitos de calidad y los atributos candidatos; determinando, a continuación, la "vista de parámetros de calidad": así, a cada elemento del esquema conceptual se le puede asociar un parámetro de calidad, por ejemplo, a la nota de selectividad la precisión y la oportunidad. Posteriormente se objetiviza los parámetros subjetivos añadiendo etiquetas a los atributos del esquema conceptual (la fuente y la fecha de las notas de selectividad) y se integran las diferentes vistas.

3 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). El modelo de datos es la base para todo el trabajo de diseño posterior, y es el principal factor determinante de la calidad del diseño del sistema global. (Meyer, 1988; Sager, 1988).

Desarrollar modelos de datos de mejor calidad será el paso principal hacia cons-

truir sistemas de mejor calidad y por consiguiente bases de datos de mejor calidad. Bubenko (1986) mencionó la falta total de una guía para evaluar los modelos conceptuales en las metodologías existentes de diseño de los sistemas de información.

ALUMNO	NOTA SELECTIVIDAD	NOTA MEDIA CARRERA
Willian Smith	8 <30/10/90, MEC>	8 <30/10/90, MEC>
Gene Hackman	9 <30/10/90, MEC>	9 <30/10/90, MEC>
...

"El modelo de datos es la base para todo el trabajo de diseño posterior, y el principal factor determinante de la calidad del diseño del sistema global"

Lo que es necesario es un marco de evaluación dentro del cual, dos modelos cualesquiera, no importa cuan diferentes sean, puedan ser comparados de forma precisa, objetiva y comprensiva. Todo esto nos lleva a pensar en la necesidad de abordar seriamente el tema de la calidad de los modelos conceptuales.

Autores	Propiedades
Batini et al. (1992)	compleción, corrección, minimalidad, expresividad, legibilidad, autoexplicación, extensibilidad y normalidad.
Reingruber M. y Gregory W. (1994)	corrección conceptual, compleción conceptual, corrección sintáctica, compleción sintáctica, conocimiento de la empresa.
Boman et al. (1997)	Facilidad de comprensión, corrección semántica, estabilidad, compleción, enfoque conceptual.

TABLA 2 • Propiedades "deseables" para el modelo conceptual de datos

Algunos autores han definido la calidad de los modelos conceptuales como una lista de propiedades "deseables" que debe satisfacer el modelo de datos (véase tabla 2) 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, se mezclan

características de la especificación con las del método y del lenguaje de modelado, se presupone la existencia de diseño/implementación, y se presentan objetivos poco realistas o imposibles de alcanzar (Lindland et al., 1994).

Una estrategia más adecuada para atacar la calidad es definir marcos de referencia que organicen y estructuren los conceptos claves y características en el modelado de datos.

Recientemente, se han publicado algunos marcos de referencia que permiten abordar la calidad en el modelado conceptual de una manera más sistemática (Moody y Shanks, 1994; Moody et al., 1998; Shanks y Darke, 1997; Lindland et al., 1994; Krogstie et al., 1995; Eick, 1991; Kesh, 1995; Schuette y Rotthowe, 1998). A continuación analizaremos cada uno de ellos.

3.1 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

DIVULGACIÓN

lo que se pueden identificar los siguientes elementos (véase figura 2):

- Audiencia (A): 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 (M): conjunto de todas las sentencias expresadas explícita o implícitamente.

- Lenguaje (L): conjunto de todas las sentencias que se pueden expresar de acuerdo al vocabulario y la gramática de los "lenguajes de modelado" (modelos) utilizados.

- Dominio (D): conjunto de todas las sentencias que serían correctas y relevantes acerca del problema.

- Interpretación de la audiencia (I): conjunto de todas las sentencias de las que la audiencia piensa que consta el modelo.

- Conocimiento de los participantes (K): unión de los conjuntos de sentencias de todos los actores sociales individuales.

Como es natural, consideraremos que un esquema tiene una mayor calidad semántica cuanto mejor sea la correspondencia entre el esquema conceptual que hemos diseñado y el dominio que pretendemos 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

medios para mejorar los diferentes tipos de calidad. La tabla 3 muestra los objetivos y los medios para alcanzarlos de acuerdo los conceptos de calidad propuestos en los marcos de Lindland et al.(1994) y Krogstie et al. (1995), que si bien unos son extensiones de otros, persiguen el mismo objetivo, que es entender el concepto de la calidad en los modelado conceptual basándose en conceptos semióticos.

Aunque el marco presentado y sus posteriores extensiones, son realmente una contribución para tener un mejor entendimiento de los aspectos de la calidad concernientes al modelado conceptual, tienen un alto nivel de abstracción y no son realmente aplicables para los profesionales.

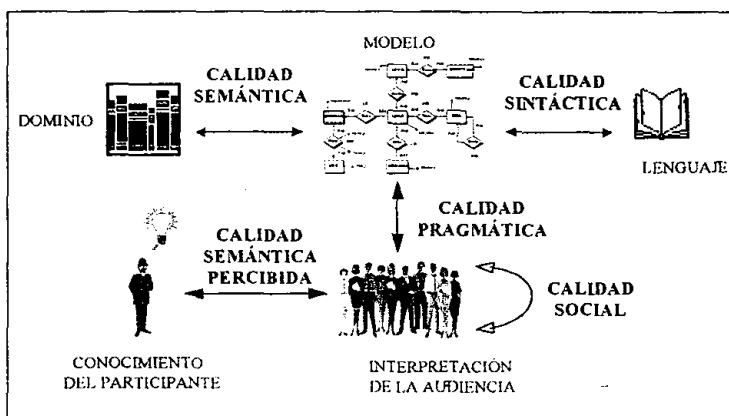


FIGURA 2 • Marco para la calidad del modelado conceptual, Krogstie et al. (1995)

el problema) como la compleción (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.

Este marco propone además distintos

3.2 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):

- 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

Tipos de calidad	Objetivos	Medios	
		Propiedades modelo	Actividades
SINTÁCTICA	Corrección sintáctica	Sintaxis formal	Verif. Sintáctica
SEMÁNTICA	Validez viable Compleción viable	Semántica formal Modificabilidad	Verif. Consistencia Inserción sentencias Borrado sentencias
SEMÁNTICA PERCIBIDA	Validez viable Compleción viable		Inserción sentencias Borrado sentencias Entrenamiento
PRAGMÁTICA	Comprensión viable	Economía expresiva Estética	Inspección Visualización Filtrado Presentación diag. Paráfrasear Explicación Entrenamiento Ejecución Animación Simulación
SOCIAL	Acuerdo viable	Ejecutabilidad Modelado conflicto	Análisis punto vista Resolución conflicto Fusión de modelos

TABLA 3 • Objetivos y medios para conseguir la calidad (Krogstie et al., 1995)

DIVULGACIÓN

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.

Para Moody estos factores significan:

- **Compleción:** capacidad del modelo de tener toda la información necesaria para cumplir los requisitos del usuario.
- **Integridad:** grado en el que las reglas del negocio que se aplican a los datos están definidas en el esquema conceptual.
- **Flexibilidad:** facilidad con la que el esquema se puede adaptar a los cambios en los requisitos.
- **Comprensibilidad:** facilidad con la que el esquema puede ser entendido.
- **Corrección:** se refiere a si el esquema cumple las reglas de las técnicas de modelado utilizadas.
- **Simplicidad:** significa que el esquema contiene los mínimos constructores posibles.
- **Integración:** nivel de consistencia del esquema con el resto de los datos de la organización.

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, como muestra la tabla 4.

Moody (1998) propone 25 métricas clasificadas según los factores de calidad de la figura 4; algunos calculados de forma objetiva (p. ej. número de entidades del

esquema), mientras que otros resultan de la puntuación subjetiva de los implicados en el diseño (véase tabla 5).

3.3. PROPUESTA DE SHANKS Y DARKE

Shanks y Darke (1997) han propuesto integrar el marco de Moody y Shanks (1994) con el de la universidad 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.

Los conceptos comunes son integrados en este marco, y los conceptos disjuntos son incorporados al mismo, como se muestra en la figura 5. En dicha figura se indican los conceptos referentes a la calidad en el modelado de datos basados en la teoría, y además los conceptos que soportan la evaluación de la calidad en la práctica. Los conceptos que están basados tanto en la práctica como en la teoría se muestran en la parte central (sombreada con un color más oscuro) de la figura 5, la cual permite entender la relación entre ambos marcos y como uno informa al otro. Este marco integrado puede aplicarse en cualquier etapa del proceso de modelado conceptual, y tiene en cuenta el concepto de calidad tanto en el modelado del producto como en el proceso de modelado del mismo.

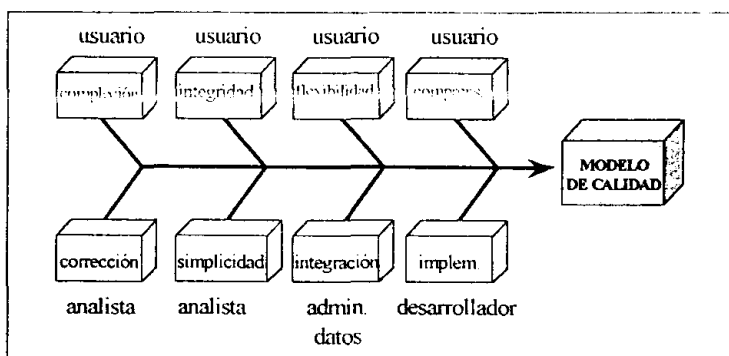
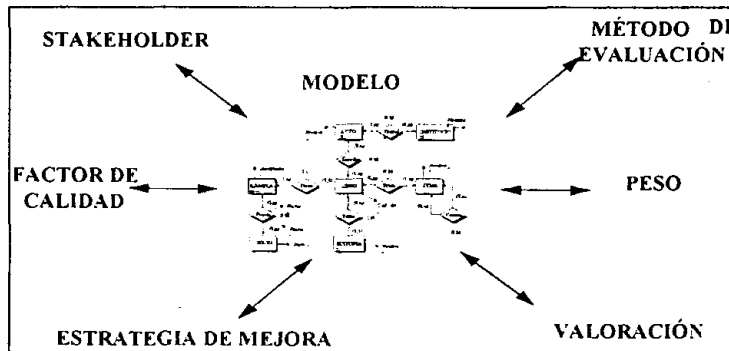


FIGURA 4 • Factores de calidad, (Moody et al., 1998)

	COMPREN.	SIMPLIC.	FLEXIB.	COMPLEC.	IMPLEM.	INTEGR.
COMPREN.						
SIMPLIC.	+					
FLEXIB.		+				
COMPLEC.	+					
IMPLEM.		+	-	-		
INTEGR.	-		+	-	+	

TABLA 4 • Interacciones positivas y negativas entre factores

DIVULGACIÓN

FACTOR DE CALIDAD	MÉTRICAS
Compleción	<ul style="list-style-type: none"> Nº de elementos del modelo de datos que no corresponden con requisitos de usuario Nº de requisitos de usuario no representados en el modelo de datos Nº de elementos de datos que corresponden a requisitos de usuario pero definidos de forma inexacta Nº de inconsistencias con el modelo de procesos
Integridad	<ul style="list-style-type: none"> Nº de reglas del negocio que no se hacen cumplir por el modelo de datos Nº de restricciones de integridad incluidas en el modelo de datos que no corresponden a políticas del negocio
Flexibilidad	<ul style="list-style-type: none"> Nº 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"> Nº de violaciones de las convenciones de modelado de datos Nº de violaciones a las formas normales Nº de instancias de redundancia en el modelo
Simplicidad	<ul style="list-style-type: none"> Nº de entidades Nº de entidades e interrelaciones Nº de constructores (aNE + bNR + cNA)
Integración	<ul style="list-style-type: none"> Nº de conflictos con el modelo de datos corporativo Nº 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 Nº de elementos físicos incluidos en el modelo de datos

TABLA 5 • Métricas para evaluar la calidad de modelos EIR (Moody, 1998)

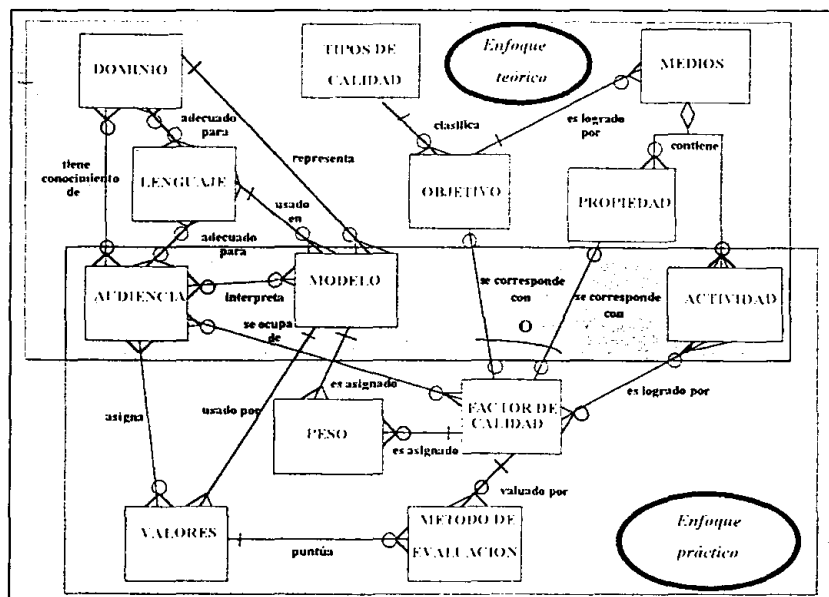


FIGURA 5 • Metamodelo propuesto por Shanks y Darke (1997)

3.4 PROPUESTA DE KESH

Kesh (1995) propuso el desarrollo de un modelo, una metodología y métricas para evaluar la calidad del modelo E/R.

El desarrollo del modelo está basado en conceptos tomados de dos disciplinas, que son la inteligencia artificial (específicamente de los sistemas basados en el conocimiento) y la ingeniería de software. Se basó en la inteligencia artificial porque los aspectos referidos a la validación y verificación de los sistemas basados en el conocimiento (y por consiguiente la calidad de los sistemas) ha sido tratada especialmente en la literatura referida a la inteligencia artificial. Desarrollar un sistema basado en el conocimiento y un modelo E/R es similar en el sentido de que ambos requieren de la captura de información (conocimiento) a cerca de un dominio, proporcionada por los usuarios (expertos), y tal proceso es un arte más que una ciencia exacta.

Además Kesh (1995) tomó como referencia a la ingeniería de software, porque es una disciplina que tiene literatura muy rica sobre calidad de software, métricas, etc...

La figura 6 (página siguiente), muestra el marco de referencia utilizado para evaluar la calidad, que considera dos tipos de componentes: de comportamiento y ontológicos.

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 y el contenido se refiere a los atributos de las entidades.

Factores que influyen en la calidad de la estructura:

- Adecuación al problema: la adecuación de la estructura, se refiere a si el diseño del modelo E/R refleja la estructura del problema que se está modelando.
- Validez: se refiere al cumplimiento de principios técnicos de diseño.
- Consistencia: se refiere a si el modelo no muestra contradicciones.
- Concisión: se refiere a que el modelo E/R no tenga redundancia.

Factores que influyen en la calidad del contenido:

- Compleción: se refiere a que todos los atributos que sean relevantes deben ser incluidos.
- Cohesión: se refiere a la cercanía que existe entre los atributos de una entidad.

DIVULGACIÓN

- Validez: se refiere a que los atributos sean correctamente asignados a las entidades.

Como muestra la figura 6, existe una fuerte relación entre la estructura y el contenido. Para obtener un modelo E/R de buena calidad, es necesario evaluar la calidad tanto del contenido como de la estructura.

Los componentes ontológicos, la estructura y el contenido, determinan el comportamiento del sistema.

Factores que influyen en la calidad del comportamiento:

- Usabilidad: la usabilidad se refiere a si un diagrama E/R es conveniente y práctico para su uso. Los diseñadores y los usuarios tienen diferentes puntos de vista a cerca de la usabilidad, por eso es conveniente subdividirla en usabilidad (usuario) y usabilidad (diseñador).

- Mantenibilidad: es la facilidad con que un diagrama E/R puede ser modificado, corregido y extendido ante futuros requerimientos.

- Precisión: se refiere a la exactitud con que el diagrama E/R refleja el problema que se esta modelando.

- Rendimiento: se refiere aun eficiente diseño del diagrama E/R, y la eficiencia es el número de entidades, relaciones y atributos con relación al tipo de tarea que la base de datos representa.

La tabla 7 muestra como los factores ontológicos afectan a los factores de comportamiento.

Además Kesh (1995) propone una metodología y métricas para evaluar la calidad de los modelos E/R que detallamos a continuación.

Una vez construido un modelo 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 combi-

nar los valores de los componentes ontológicos que le son relevantes.

- 3) Combinar los valores de los componentes de comportamiento para calcular el valor global de la calidad.

- 4) Si el valor de un componente ontoló-

dología para el diseño y transformación de esquemas conceptuales, denominada ANNAPURNA, que ha sido implementada en una herramienta CASE.

Dentro de la metodología Eick (1991) ha propuesto una métrica para evaluar la calidad de los esquemas conceptuales.

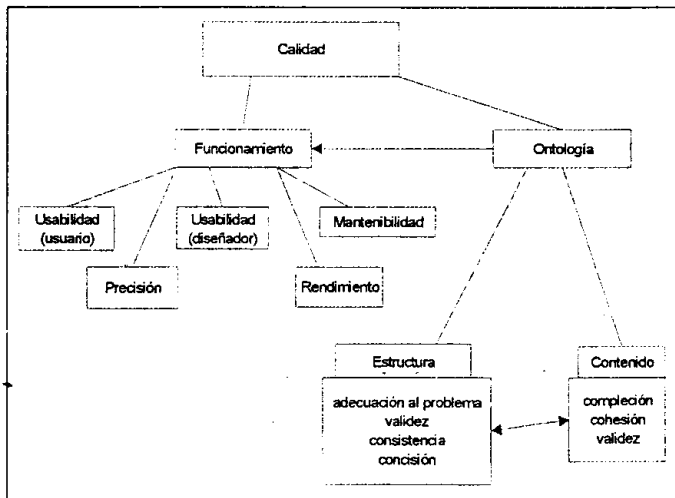


FIGURA 6 • Marco de Referencia para la evaluación de la calidad (Kesh, 1995)

Los esquemas conceptuales planteados por Eick (1991) se denominan "diagramas S" basados en el modelo relacional binario y el SDM (Semantic Database Model). Estos diagramas soportan los siguientes conceptos básicos:

- clases
- conexiones entre clases
- atributos
- jerarquías de generalización (superclases/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

Una vez diseñados los diagramas S se evalúan utilizando las siguientes medidas: 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:

$$f(qu_1, qu_2, qu_3, qu_4, qu_5)$$

3.5. PROPUESTA DE EICK

Eick (1991) ha propuesto una meto-

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

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

TABLA 7 • Modelo para evaluar la calidad de los modelos E/R

Se dice que S_j es mejor que S_i , si f le asigna un valor más alto a S_j que a S_i y S_j es válido.

La función f se define de la siguiente manera:

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

Eick (1991) propone dos algoritmos para poder decidir si una dependencia funcional o en existencia presente en el universo de discurso, se expresa en un diagrama S.

Esta métrica se encuentra automatizada e integrada a la metodología, la cual le ofrece a los diseñadores medidas que pueden facilitar la labor de transformación y refinamiento de los esquemas conceptuales y en consecuencia mejorar su calidad.

3.6 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 tal subjetividad debe ser manejada y comprendida.

El énfasis de la subjetividad en el proceso de modelado requiere medidas específicas para el manejo de la subjetividad para poder desarrollar modelos subjetivamente comparables y probables. El hecho de resaltar la subjetividad enfatiza la importancia del diseñador.

Como la subjetividad del modelado no puede ser eliminada completamente pero solo controlada, es necesario contar con reglas que tengan que ser seguidas en el proceso de modelado. 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. Además, un usuario del modelo, quien no estuvo involucrado en el proceso de modelado, podría entender los modelos, solo si conoce las reglas subyacentes del proceso de modelado.

Schuette y Rotthowe (1998) 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.

■ La arquitectura GoM, que es un mar-

Principios	Objetivos
Principio de adecuación de la construcción	<ul style="list-style-type: none"> • Consenso a cerca de la definición de la definición del problema • Consenso a cerca de la representación del modelo <ul style="list-style-type: none"> • Consistencia intra-modelo • Consistencia inter-modelo • Minimalidad
Principio de adecuación del lenguaje	<ul style="list-style-type: none"> • Corrección del lenguaje • Adaptación del lenguaje <ul style="list-style-type: none"> • Poder semántico • Formalización • Comprensibilidad del lenguaje
Principio de la eficiencia económica	<ul style="list-style-type: none"> • Consenso • La comprensibilidad y aplicación del lenguaje • Comparabilidad estructura sistemática
Principio de claridad	<ul style="list-style-type: none"> • Diseño jerárquico • Diseño del esquema <ul style="list-style-type: none"> • Filtrado • Filtros metódicos • Filtros de contenido
Principio del diseño sistemático	<ul style="list-style-type: none"> • Consistencia inter-modelo entre los modelos de la estructura y el comportamiento • Arquitecturas de los sistemas de información
Principio de comparabilidad	<ul style="list-style-type: none"> • Comparabilidad a nivel de meta modelo • Transformación completa • Traducción consistente • Comparabilidad a nivel del modelo

TABLA 8 • Principios de la Guía de Modelado (GoM)

co estructural en el cual son ubicados cada uno de los componentes de esta guía.

La idea de los autores al proponer una Guía de Modelado (GoM) es presentar un marco de principios que mejore la calidad de los modelos de información reduciendo el subjetivismo en el proceso de modelado de la información. La calidad de los modelos se basa en recomendaciones para lograra un eficiente, comprensivo y correcto diseño de los modelos de información.

La tabla 8 muestra los objetivos de cada uno de los seis principios generales de GoM.

Los principios enumerados en la tabla 8 forman la base de la Arquitectura GoM.

La Guía del Modelado, además del esfuerzo de estandarización, tiene en cuenta la interrelación entre el problema a ser modelado y su representación, la eficiencia económica del modelado y medidas para mejorar la claridad del modelo.

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 creemos han tenido una mayor repercusión. Si bien todas estas propuestas han significado un gran contribución, creemos que es necesaria una mayor investigación en los aspectos rela-

cionados 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 en 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, ya sea el proceso de modelado como el proceso de obtención y carga de los datos, ya que la mayor parte de los estudios revisados en este artículo se centran en la calidad del producto (modelo de datos conceptual y datos). ■