



# VI jornadas Ingeniería del Software y Bases de Datos

Almagro (Ciudad Real)  
21 - 23 de Noviembre de 2001



Grupo Alarcos  
Departamento de Informática  
Escuela Superior de Informática

Editores:  
Oscar Díaz  
Arantza Illarramendi  
Mario Prattini

# jisbd 2001

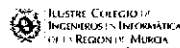
## PATROCINAN



**soluziona** servicios profesionales



## COLABORAN

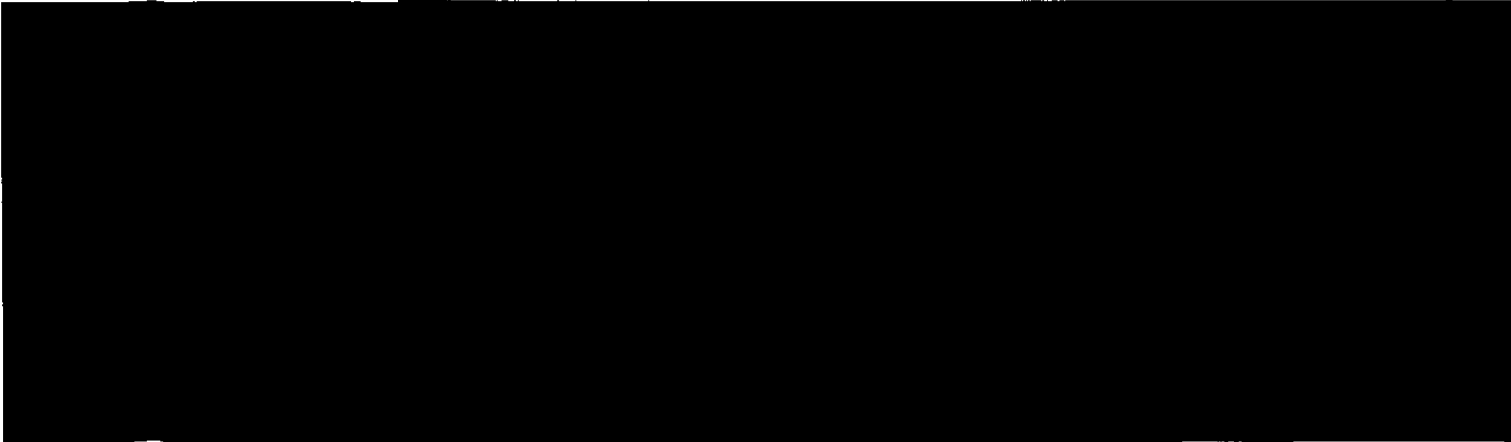


EIEO  
Instituto de Estudios e Investigación Científica  
CORE  
Colegio Oficial de Ingenieros en Informática de Murcia



JISBD' 2001  
VI Jornadas de Ingeniería del Software y Bases de Datos

Almagro (Ciudad Real) 21 - 23 de Noviembre de 2001



jisbd  
2001

# JISBD' 2001

## Jornadas de Ingeniería del Software y Bases de Datos

21 y 23 de Noviembre de 2001  
Almagro (Ciudad Real)



Grupo Alarcos

Departamento de Informática

Escuela Superior de Informática

### Editores

Oscar Díaz  
Arantza Illaramendi  
Mario Piattini

**Presidencia**

Óscar Díaz  
Arantza Illarramendi  
Universidad del País Vasco

**Editores**

Oscar Díaz  
Arantza Illarramendi  
Mario Piattini

**Organizadas por:**

Grupo Alarcos

Departamento de Informática

Escuela Superior de Informática



José Francisco Aldana	<i>Universidad de Málaga</i>
Idoia Alarcón	<i>Universidad Autónoma de Madrid</i>
João Araújo	<i>Universidade Nova de Lisboa</i>
Manuel Barrena	<i>Universidad de Extremadura</i>
Pere Botella	<i>Universidad Politécnica de Cataluña</i>
Hilario Canos	<i>Universidad Politécnica de Valencia</i>
Matilde Celma	<i>Universidad Politécnica de Valencia</i>
João A. Correia Lopes	<i>Universidade do Porto</i>
Dolors Costal	<i>Universidad Politécnica de Cataluña</i>
Carmen Costilla	<i>Universidad Politécnica de Madrid</i>
Juan Manuel Cueva	<i>Universidad de Oviedo</i>
Pablo de la Fuente	<i>Universidad de Valladolid</i>
Carlos Delgado Kloos	<i>Universidad Carlos III</i>
Xavier Franch	<i>Universidad Politécnica de Cataluña</i>
Juan Garbajosa	<i>Universidad Politécnica de Madrid</i>
Jesus García Molina	<i>Universidad de Murcia</i>
Alfredo Goñi	<i>Universidad del País Vasco</i>
Juan Hernández	<i>Universidad de Extremadura</i>
Arturo Jaime	<i>Universidad del País Vasco</i>
Natalia Juristo	<i>Universidad Politécnica de Madrid</i>
Esperanza Marcos	<i>Universidad Rey Juan Carlos</i>
José Manuel Marqués	<i>Universidad de Valladolid</i>
Eduardo Mena	<i>Universidad de Zaragoza</i>
Roberto Moriyón	<i>Universidad Politécnica de Madrid</i>
Juan José Moreno	<i>Universidad Autónoma de Madrid</i>
José Parcts	<i>Universidad de Granada</i>
Oscar Pastor	<i>Universidad Politécnica de Valencia</i>
Ernesto Pimentel	<i>Universidad de Málaga</i>
António Rito Silva	<i>Universidade Técnica de Lisboa</i>
Nieves R. Brisaboa	<i>Universidad de La Coruña</i>
Félix Saltor	<i>Universidad Politécnica de Cataluña</i>
José Samos	<i>Universidad de Granada</i>
Ernest Teniente	<i>Universidad Politécnica de Cataluña</i>
Ambrosio Tova	<i>Universidad de Murcia</i>
Jose María Troya	<i>Universidad de Málaga</i>
Antonio Vallecillo	<i>Universidad de Málaga</i>
Juan Carlos Yelmo	<i>Universidad Politécnica de Madrid</i>

© Los autores

Primera edición, 2001

I.S.B.N.: 84-699-6275-2

Depósito Legal: CR-506/2001

Imprime: GRAFICAS GARRIDO, S.L.

## Colaboradores

Alberto Abelló	Luis Llopis Torres
Celeste Campo Vázquez	Julio Mariño Carballo
Carlos Canal	Enric Mayol
José Angel Carsí Cubel	Pedro Merino
José María Cavero Barca	Nathalie Moreno Vergara
Rodrigo Cerón	Juan Manuel Murillo
Manuel Collado Machuca	Antonio Jesús Nebro Urbaneja
Yania Crespo González-Carvajal	Abelardo Pardo Sánchez
Antonio César Gómez Lora	Joan Antoni Pastor
Cecilia Delgado	Vicente Pelechano
Francisco Durán	Carne Quer
José Fernán Martínez	Maria Ribera Sancho
José Luis Fernández Alemán	Elena Rodríguez
Lidia Fuentes Fernández	María del Mar Roldán García
Francisco José Galán Morillo	Joaquín Nicolás Ros
María del Mar Gallardo Melgarejo	Bartolomé Rubio
José Jesús García Rueda	José Sáez Martínez
Eladio Garví	Juan Sanchez
Miguel Goulão	María Isabel Sánchez-Segura
Ángel Herranz Nieva	Fernando Sánchez Figueroa
José Ramón Hoyos Barceló	Almudena Sierra Alonso
Emilio Insfran	Toni Urpí
Vicente Luque	Jesús Villamor Lugo
	María Inmaculada Yague del Valle

## Tutoriales

Toni Urpí  
Miguel Toro

Universidad Politécnica de Cataluña  
Universidad de Sevilla

## Ponencias Invitadas

Isidro Ramos  
Antoni Olivé

Universidad Politécnica de Valencia  
Universidad Politécnica de Cataluña

## **Comité Organizador**

### **Presidente**

Mario Piattini  
Grupo Alarcos - UCLM

Ismael Caballero	Universidad de Castilla-La Mancha
Coral Calero	Universidad de Castilla-La Mancha
José Antonio Cruz	Universidad de Castilla-La Mancha
Eduardo Fernández-Medina	Universidad de Castilla-La Mancha
Félix Óscar García	Universidad de Castilla-La Mancha
Marcela Genero	Universidad de Castilla-La Mancha
Mar Jiménez	Universidad de Castilla-La Mancha
José Manuel Marqués	Universidad de Valladolid
Luis Márquez	Universidad de Castilla-La Mancha
M <sup>a</sup> Dolores Mateos	Universidad de Castilla-La Mancha
Ana Moreira	Universidade Nova de Lisboa
Macario Polo	Universidad de Castilla-La Mancha
Francisco Ruiz	Universidad de Castilla-La Mancha
Manuel Ángel Serrano	Universidad de Castilla-La Mancha

## Prólogo del Comité de Programa

Este volumen recoge los trabajos presentados en las VI Jornadas de Ingeniería de Software y Bases de Datos celebrados en Almagro los días 21, 22 y 23 de noviembre de 2001. Esta edición ha buscado cohesionar más si cabe, a las comunidades de Bases de Datos e Ingeniería de Software, estableciendo un único comité de programa, y esforzándose por buscar talleres y conferenciantes de interés común. Todo ello con el objetivo de fomentar la creciente necesidad de equipos multi-disciplinares que puedan afrontar los cada vez más exigentes desafíos de la sociedad de la información.

Para participar en las sesiones técnicas se recibieron un número record de trabajos: 89. Este número avala, por un lado, la creciente madurez del área, y por otro, el prestigio de las Jornadas, fruto del buen hacer de las ediciones anteriores. Finalmente, se aceptaron 38 trabajos.

A continuación facilitamos algunos datos sobre la evaluación, ya que velar por la calidad del proceso de revisión es nuestra labor principal como presidentes del comité de programa. La práctica totalidad de los trabajos fueron revisados por tres miembros del comité de programa. El evaluador tenía que definirse como "experto", "conocedor" o "poco familiarizado" con la temática del trabajo, siendo cuatro los posibles resultados: "rechazar", "posiblemente rechazar", "posiblemente aceptar" y "aceptar". El 82% de los trabajos contó con un evaluador "experto". De los trabajos aceptados, ninguno de sus revisores "expertos" decidió rechazarlo, el 37% contó con un "experto" que concluyó con "aceptar", y en el 87% de los casos NO hubo disparidad de opiniones, es decir los tres revisores concluyeron con un "aceptar" o un "posiblemente aceptar". De los trabajos rechazados, el 58% tuvo un evaluador "experto" que concluyó con un "rechazar".

Pero además de las sesiones técnicas, esta edición ha apostado por los talleres y tutoriales. Los talleres promovidos por los comités respectivos, favorecen un intercambio amplio y ágil, tanto de ideas ya asentadas, como de barruntos o intuiciones que en un entorno más formal difícilmente saldrían a la luz. Con esta finalidad, se organizaron seis talleres, a saber: Impacto de la tecnología XML en las bases de datos, Ingeniería del software basada en componentes distribuidos, Almacenes de datos y tecnología OLAP, Ingeniería del Software orientada a la Web, Desarrollo de software preciso y Evolución del software.

Contamos también con una mesa redonda y tres conferenciantes invitados. La primera trató el tema de "*La investigación en Ingeniería del Software*" y fue coordinada el Dr. Pere Botella. Respecto a los conferenciantes, se contó con la presencia de F. Arbab que habló sobre Panta Rei, con P. Fankhaus que trató sobre XQuery y una tercera conferencia que, en el momento de escribir este prólogo aún quedaba por determinar.

Ya para finalizar con el programa, esta edición ha introducido dos novedades: la sesión dedicada a trabajos ya publicados, y los tutoriales. La primera nace con el deseo de presentar en las Jornadas aquellos trabajos nacionales que, habiendo sido ya publicados en revistas de prestigio, puedan servir de referencia a otros grupos por su



calidad. El paisanaje con los ponentes puede así favorecer cierto espíritu emulador que acreciente la presencia de los grupos nacionales en los foros internacionales. En este apartado se seleccionaron los trabajos de Jaime Gómez et al. "Conceptual modeling of device-independent Web applications" y de María Paloma Díaz et al. "Laberinth: un modelo formal para la especificación de sistemas hipermedia". Respecto a los tutoriales, se ofrecieron los siguientes: "Introducción a las bibliotecas digitales multimedia" y "Modelado conceptual de aplicaciones Web".

Obviamente, estas Jornadas no hubieran sido posible sin la participación de muchas personas. En primer lugar, agradecer a la comunidad científica su apoyo a estas Jornadas enviándonos sus trabajos. Asimismo, nuestra gratitud hacia los miembros de los diferentes Comités de Programa por la cuidadosa revisión de los trabajos enviados. Por último, nuestro más sincero reconocimiento a todos los miembros del Comité Organizador, por su disponibilidad y entrega durante todo este tiempo.

Donostia, octubre de 2001

Oscar Díaz & Arantza Illarramendi  
Presidentes del Comité Organizador

## Prólogo del Comité Organizador

Desde 1996 se vienen celebrando en España las Jornadas de Ingeniería del Software y las de Investigación y Docencia en Bases de Datos. En 1999 ambos eventos se unieron en las IV Jornadas de Ingeniería de Software y Bases de Datos, celebradas en Cáceres, y que constituyeron un éxito tanto desde el punto de vista científico como del de organización, reconfirmándose este éxito en las últimas Jornadas celebradas el año pasado en Valladolid.

Sin duda, las JISBD se han consolidado como el foro de encuentro y discusión para el intercambio de experiencias y conocimiento entre los principales investigadores de la Ingeniería del Software y de las Bases de Datos, como lo demuestra el casi centenar de comunicaciones enviadas en esta edición.

JISBD 2001 se desarrolla en ocho sesiones técnicas, dos ponencias invitadas, tres conferencias, un panel de expertos y seis talleres de trabajo (que incluyen el impacto de la tecnología XML en las bases de datos, la Ingeniería del Software basada en componentes distribuidos, los almacenes de datos y la tecnología OLAP, la Ingeniería del Software orientada a la Web, el desarrollo de software preciso y la evolución del software). Además, se incluyen dos tutoriales: "Introducción a las Bibliotecas Digitales Multimedia" y "OOWS: Una Aproximación para el Modelado Conceptual", que unidos a otros tres que se celebran en conjunción con las JISBD: "Data Security", "XML and Databases" y "Evaluación y Mejora de Procesos Software", ofrecen una formación avanzada en estas áreas.

A todas estas actividades se ha unido la celebración de las II Jornadas de Bibliotecas Digitales (JBIDI), las I Jornadas de Programación y Lenguajes (PROLE), las I Jornadas de Sistemas de Información Geográfica (JSIG), y la Jornada de Seguimiento de proyectos CICYT, así como otras reuniones y seminarios: ADIS, RETISBD, ZOCO, etc. que contribuyen a que Almagro se convierta en esta semana de noviembre en la "capital de la informática".

Resulta imprescindible aprovechar estas líneas para agradecer a los miembros del Comité de Programa y especialmente a sus presidentes, la Dra. Arantza Illarramendi y el Dr. Óscar Díaz, el esfuerzo realizado durante todo este tiempo. Igualmente es necesario destacar la labor realizada tanto por los miembros de los comités de tutoriales y ponencias invitadas, como por los coordinadores y comités de programas de los diferentes talleres de trabajo.

Asimismo, quiero agradecer el esfuerzo realizado a todos los miembros del Comité Organizador, que este año se han enfrentado a una tarea particularmente difícil, debido a la gran cantidad de eventos que han coincidido con las JISBD; su entrega y dedicación han hecho posible la celebración de estas jornadas que esperamos resulten del interés de todos los que en ellas participan.

Mario Piattini Velhuis  
Presidente del Comité Organizador

## Índice

<b>Conferencias .....</b>	<b>1</b>
XQuery between data and documents: bridging the gap..... <i>Peter Fankhauser</i>	3
Panta Rei (Everything Flows): A Calculus of Channels..... <i>Farhud Arbab</i>	5
<b>Ponencias Invitadas .....</b>	<b>7</b>
Conceptual Modeling of Device-Independent Web Applications..... <i>Juime Gómez, Cristina Cachero, Oscar Pastor</i>	9
LABYRINTH: Un modelo formal para la especificación de sistemas hipermedia..... <i>Paloma Diaz</i>	35
<b>Sesiones .....</b>	<b>51</b>
<b>Sistemas Distribuidos.....</b>	<b>53</b>
Semicomposición Distribuida..... <i>J. F. Aldana, A.C. Gómez, M.M. Roldán</i>	55
CORBA Lightweight Components: An Early Report..... <i>Diego Sevilla, José M. García, Antonio Gómez</i>	69
JReplica: Modelo de replicación transparente e independiente del ORB... <i>José Luis Herrero, Fernando Sánchez, Miguel Toro</i>	85
Un Estilo Arquitectónico para la Federación de Sistemas Basados en Procesos..... <i>Angel Martínez, José H. Canós, Jesús D. García-Consuegra</i>	99
Orientando a Aspectos la captura de datos de Internet para activar Reglas de Decisión en Agentes Web..... <i>Rafael Corchuelo, Jesús S. Aguilar, J.L. Arjona, Miguel Toro y José Riquelme</i>	113
<b>Pruebas y Mantenimiento.....</b>	<b>129</b>
SEGESOFT: Entorno de Entrenamiento para la Gestión de Proyectos Software..... <i>J. C. Riquelme, I. Ramos, J. Aguilar-Ruiz, F. Ferrer, M. Toro, J.J. Dolado, A. Ruíz de Infante, J. Tuyá, P. Fernández, M.A. Prieto, M. Ruíz-Carreira, D. Rodríguez-García, M. Satpathy, R. Harrison, R. Matilla, M.A. Álvarez</i>	131
Cómo Modelar El Efecto En La Planificación De Proyectos Por La Dctención Tardía De Defectos; La Red De Calidad Incremental..... <i>Sergio Coronado, José Alberto Jaén</i>	145

Entorno Global para la Gestión del Proceso de Mantenimiento del Software.....	157	Verificación Automatizada Eficiente de Sistemas Concurrentes Especificando sus Propiedades con Fórmulas de una Lógica de Intervalos.....	375
<i>Francisco Ruiz, Mario Piattini</i>		<i>Miguel J. Hornos, Manuel I. Capel</i>	
Esquema de Caracterización para la Selección de Técnicas de Pruebas.....	171	Generación de Restricciones del Entorno para Verificación Modular Utilizando Reglas de Asociación.....	391
<i>Oscar Dieste Tubío, Sira Vegas Hernández</i>		<i>Claudio de la Riva, Javier Tuya y José R. de Diego</i>	
Entorno basado en el uso de técnicas formales para el desarrollo y prueba de sistemas.....	189	Verifying Reuse Contracts with a Component Model.....	405
<i>Marisol Sánchez-Alonso, José Luis Herrero</i>		<i>Agustín Cernuda del Río, Jose Emilio Labra Gayo, Juan Manuel Cueva Lovelle</i>	
<b>Miscelánea.....</b>	<b>205</b>	From Maude Specifications to SOAP Distributed Implementations: A Smooth Transition.....	419
Selección Conjunta de Componentes COTS.....	207	<i>Antonio Albarrán, Francisco Durán, Antonio Vallecillo</i>	
<i>X. Burgués, C. Estay, X. Franch, J.A. Pastor, C. Quer</i>		<b>Calidad Software.....</b>	<b>435</b>
Un Entorno para el Desarrollo de Modelos de Flujo de Trabajo.....	223	Una experiencia en la Obtención de un Instrumento Fiable para Medir Satisfacción de Clientes de Software.....	437
<i>M<sup>a</sup> Carmen Penadés, José H. Canós</i>		<i>María Inés Lund, Belén Forcada, Sergio Zapata, Miriam Herrera</i>	
Representación de Restricciones Difusas en un Modelo Conceptual.....	237	Evaluación de la Comprensión Semántica del Modelado Dinámico en UML.....	449
<i>Angélica Urrutía, José Galindo, Mario Piattini</i>		<i>Mari Carmen Otero, José Javier Dolado</i>	
El rendimiento en Sistemas de Programación Persistentes.....	253	Sistema de Gestión de Meta-información para la Integración de Métricas de Calidad de Modelos UML.....	465
<i>J. Baltasar García Perez-Schofield, Tim B. Cooper, Emilio García Roselló, Manuel Pérez Cota</i>		<i>María N. Moreno García, Francisco J. García Peñalvo, M. José Polo Martín</i>	
Un modelo de Madurez para Investigación-Acción en Sistemas de Información.....	265	Un Método para Medir el Tamaño Funcional y Evaluar la Calidad de Sitios Web.....	477
<i>Christian A. Estay-Niculcar, Joan A. Pastor-Collado</i>		<i>Silvia Mara Abrahão, Oscar Pastor, Luis Olsina, Joan J. Fonsi</i>	
<b>Desarrollo OO.....</b>	<b>283</b>	Mejora de la consistencia en la medición del tamaño funcional de productos software.....	491
Principios y patrones en el diseño orientado a objetos.....	285	<i>Ramón Asensio Monge, Francisco Sanchis Marco, Fernando Torre Cervigón</i>	
<i>Javier Garzás, Mario Piattini</i>		<b>Bases de Datos Avanzadas.....</b>	<b>505</b>
Definición de un marco de trabajo para el análisis de refactorizaciones de software.....	297	Una propuesta de extensión de los metadatos de ODMG para la definición de esquemas externos.....	507
<i>Yania Crespo, José Manuel Marqués</i>		<i>Manuel Torres y José Samos</i>	
Tratamiento de Excepciones a Nivel de Modelado Conceptual OO.....	311	Semántica de Cambios para la Evolución Dinámica de Esquemas en Diferentes Modelos Orientados a Objetos.....	523
<i>Manoli Albert, Eva Campos, Vicente Pelechano</i>		<i>Cecilia Delgado, José Samos</i>	
Hacia Una Metodología de Desarrollo Para la Construcción de Entornos Virtuales.....	327	Métricas de calidad para almacenes de datos.....	537
<i>Gonzalo Méndez Pozo, M<sup>a</sup> Isabel Sánchez Segura, Angélica de Antonio Jiménez</i>		<i>Serrano, M.A., Pascual, C., Calero, C. y Piattini, M.</i>	
Generación automática de prototipos de interface de usuario a partir de modelos de requisitos.....	343	Diseño de un lenguaje de consulta espacial para documentos XML.....	549
<i>Juan Sánchez Díaz, Oscar Pastor López, Jorge Belenguer Faguás</i>		<i>J. E. Córcoles, P. González</i>	
<b>Técnicas Formales y Verificación.....</b>	<b>359</b>		
Towards Rigorous and Effective Type Contracts.....	361		
<i>F. J. Galán Morillo and J. M. Cañete Valdeón</i>			



<b>Ingeniería Web.....</b>	<b>565</b>
Desarrollo de Aplicaciones Web: Integración de Técnicas de IC para la Captura de Requisitos.....	567
<i>A. Sierra, P. Cáceres, E. Marcos, J. E. Pérez</i>	
Contenido en la Web: un modelo de reutilización.....	581
<i>Arantza Irastorza, Arturo Jaime, Oscar Díaz</i>	
Migration of Legacy Systems to the Web.....	601
<i>Cristina Cachero, Jaime Gómez and Antonio Párraga</i>	
Especificación de Requisitos de Calidad en Sistemas Multiorganizacionales Basados en Servicios WEB.....	615
<i>Antonio Ruiz, Amador Durán, Rafael Corchuelo y Miguel Toro</i>	

---

## CONFERENCIAS

## Referencias

1. A. Abelló, M. Oliva, E. Rodríguez, F. Saltor, C. Delgado, E. Garvía and J. Samos: Conforming Operations on Object-Oriented Database Schemas. Pendiente de publicación.
2. J. Banerjee, W. Kim, H.J. Kim and H.F. Korth: Semantics and Implementation of Schema Evolution in Object-Oriented Databases. *ACM SIGMOD*, 1987, pp. 311-322.
3. W. Kim and H. T. Chou: Versions of Schema for Object-Oriented Databases. In *Proc. of the 14<sup>th</sup> Int'l Conf. on Very Large Databases*, 1988, pp. 148-159.
4. G. Nguyen and D. Rieu: Schema Evolution in Object-Oriented Database Systems. *Data & Knowledge Engineering* 4, 1989, pp. 43-67.
5. M. T. Özsu, R. J. Peters, D. Szafron, B. Irani, A. Lipka and A. Muñoz: TIGUKAT: A Uniform Behavioral Objectbase Management System. *The VLDB Journal* Vol. 4, No.3, July 1995, pp. 445-492. Special issue on persistent object systems.
6. R. J. Peters: TIGUKAT: A Uniform Behavioral Objectbase Management System. Ph. D. thesis, Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada. Available as University of Alberta Technical Report TR94-06, 1994.
7. R. J. Peters and M. T. Özsu: Reflection in a Uniform Behavioral Object Model. In *Proc. Of the 12<sup>th</sup> Int'l Conf. on Entity-Relationship Approach*, 1993, pp. 37-49.
8. R. J. Peters and M. T. Özsu: An Axiomatic Model of Dynamic Schema Evolution in Objectbase Systems. *ACM TODS* 22(1), March 1997, pp. 75-114.
9. D. Penney and J. Stein: Class Modification in the GemStone Object-Oriented DBMS. In *Proc. of the Int'l Conf. on Object-Oriented Programming: Systems, Languages and Applications*, Oct. 1987, pp. 111-117.
10. A. Skarra and S. Zdonik: Type Evolution in an Object-Oriented Database. In *Research Directions in Object-Oriented Programming*. MIT Press, 1987, pp. 393-415.
11. R. Zicari: A Framework for Schema Updates in an Object-Oriented Database System In *Building an Object-Oriented Database System: The Story of O<sub>2</sub>*, F. Bancilhon, C. Delobel and P. Kanellakis (eds), Morgan Kaufmann, 1992, pp. 146-182.

## Métricas de calidad para almacenes de datos

Serrano, M.A., Pascual, C., Calero, C. y Piattini, M.

Grupo ALARCOS

Universidad de Castilla-La Mancha (España)

e-mail: {mserrano, cpascual, ccalero, mpiattini}@inf-cr.uclm.es

## Resumen

Las organizaciones están adoptando sistemas basados en almacenes de datos (*datawarehouses*) para gestionar la información de una forma eficiente. Es fundamental poder asegurar la calidad de la información que se encuentra dentro de los almacenes de datos, debido a que se han convertido en la principal herramienta para la toma de decisiones estratégicas. La calidad de la información depende de la calidad de su representación y de la calidad del almacén de datos. Este último aspecto, incluye la calidad del modelo multidimensional. En los últimos años varios autores han propuesto algunas guías útiles para diseñar modelos multidimensionales, sin embargo es necesario utilizar indicadores más objetivos para ayudar a los diseñadores a crear almacenes de datos de calidad. En este artículo presentamos una primera propuesta de métricas para el modelo multidimensional, su validación formal y un primer experimento llevado a cabo con el fin de validarlas empíricamente.

**Palabras Clave:** Métricas, calidad, almacenes de datos.

## 1. Introducción

Actualmente las organizaciones pueden almacenar inmensas cantidades de datos obtenidas a un precio relativamente bajo, sin embargo estos datos no proporcionan información (Gadner, 1998). Para resolver este problema, las organizaciones están adoptando almacenes de datos, los cuales se definen como "una colección de datos orientados a temas, integrados y no volátiles, que soportan la gestión de la toma de decisiones" (Inmon, 1996). Los almacenes de datos se han convertido en la tendencia más importante de la informática empresarial, ya que proporciona una información muy relevante y precisa para mejorar las decisiones estratégicas. Jarke et al. (2000) vaticinan un mercado de 12 millones de dólares americanos en el campo de los almacenes de datos.

Se han propuesto diferentes ciclos de vida y técnicas para el desarrollo de los almacenes de datos (Hammergren, 1996; Kelly, 1997; Kimball et al., 1998; Cavero et al., 2000). Sin embargo, el desarrollo de un almacén de datos es una tarea difícil y llena de riesgos. Es esencial poder asegurar la calidad de la información que contiene el almacén de datos ya que éste se ha convertido en la principal herramienta para la toma de decisiones estratégicas. (English, 1999).

La calidad de la información de un almacén de datos viene determinada por la calidad tanto del sistema completo como la calidad de la presentación de los datos (véase la figura 1). De hecho, es muy importante que los datos del almacén reflejen correctamente el mundo real, pero es también muy importante que los datos sean interpretados correctamente. En la calidad del almacén de datos, al igual que una base de datos operacional (Piatini et al., 2000), se deben considerar 3 aspectos: la calidad del SGBD (Sistema Gestor de Base de Datos) relacional o multidimensional que lo soporta, la calidad del modelo de datos (tanto conceptual, lógico como físico) y la calidad de los propios datos contenidos en el almacén.

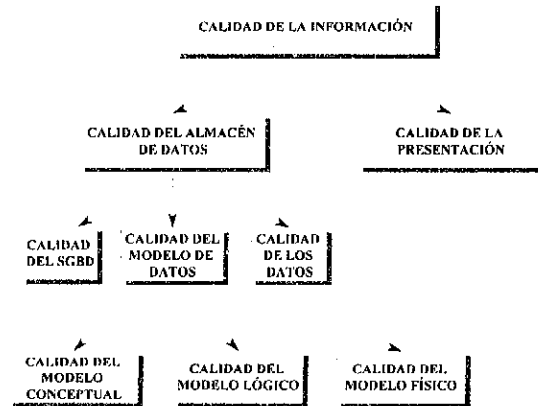


Figura 1. Calidad de la información y del almacén de datos

Para asegurar la calidad del SGBD, podemos usar un estándar internacional como el ISO 9126 (1998), o alguno de los estudios comparativos de productos existentes. Este tipo de calidad debe ser asegurada en la etapa de selección de productos dentro del ciclo de vida del data warehouse (véase, por ejemplo, Hansen y Hansen, 1997).

La calidad de los propios datos viene determinada principalmente por los procesos de extracción, filtrado, limpiado, sincronización, agregación y carga (Bouzeghoub et al., 2000; Adamson y Venerable, 1998; Kimball et al., 1998).

Como es lógico también la calidad del modelo del almacén de datos tiene una gran influencia en la calidad de la información. Modelo que puede existir a nivel conceptual -como en las propuestas de Cabibbo(1998), Golfarelli (1998;1999), Cavero et al. (2000), aunque no suele ser habitual-; a nivel lógico -para lo que se ha universalizado la utilización del "diseño en estrella", que permite buenos tiempos de respuesta y una comprensión fácil de los datos y los metadatos por parte de los usuarios y los desarrolladores (Kimball et al., 1998)-, y a nivel físico -ya que el diseñador tiene que elegir las tablas físicas, los índices y las particiones de datos que mejor representan el almacén de datos lógico y que facilitan su funcionalidad (Jarke et al., 2000; Bouzeghoub y Kedad, 2000)-.

Este trabajo se centra en la calidad del modelo de datos dimensional a nivel lógico. Varios autores han realizado recomendaciones importantes para conseguir realizar un "buen" modelo de datos dimensional (Kimball et al., 1998; Adamson and Venerable, 1998; Inmon, 1996). Sin embargo, los criterios de calidad no están lo suficientemente desarrollados como para asegurar la calidad en la práctica, ya que diferentes personas tendrán diferentes interpretaciones del mismo concepto. De acuerdo con la literatura acerca de la Gestión Total de la calidad (Total Quality Management, TQM), los criterios de medida para asegurar la calidad necesitan evitar los "argumentos de estilo" (Boman et al., 1997). El objetivo debe ser reemplazar las nociones intuitivas de "calidad" de diseño por medidas formales y cuantitativas para reducir la subjetividad y el sesgo en el periodo de evaluación. Sin embargo, para que el modelado de datos progrese desde una actividad artesanal a una disciplina ingenieril, las cualidades deseables de los modelos de datos deben hacerse explícitas (Lindland et al., 1994). Una métrica es una forma de medir un factor de calidad de una forma consistente y objetiva.

El objetivo final de nuestro trabajo es definir un conjunto de métricas para asegurar la calidad de los almacenes de datos. En la siguiente sección presentaremos un método para definir métricas correctas. Una primera propuesta de métricas para el modelo lógico de almacenes de datos se describen en la sección 3 y un ejemplo de la aplicación de las métricas propuestas se muestra en la sección 4. La sección 5 presenta la validación formal de las métricas, mientras que la sección 6 presenta una primera aproximación a los experimentos que desarrollamos para validar empíricamente estas métricas. Por último, recogemos las conclusiones y las líneas de trabajo futuro que surgen del presente artículo.

## 2. Definición de métricas válidas

La definición de las métricas debe hacerse de forma metodológica y es necesario seguir una serie de pasos para asegurar la fiabilidad de las métricas propuestas. La figura 2 muestra el método que seguimos para la proposición de métricas (Calero et al., 2001a).

En esta figura tenemos 3 actividades principales:

- **Definición de métricas.** Es el primer paso en la proposición de métricas. Esta definición está hecha de acuerdo con las características específicas del sistema que queremos medir y tiene en cuenta la experiencia de los diseñadores de esos sistemas.

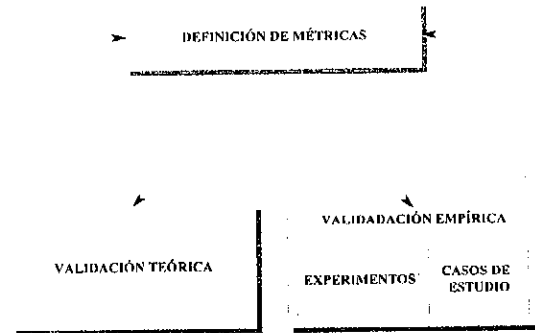


Figura 2. Pasos seguidos para la definición y validación de métricas.

- Validación teórica.** El segundo paso es la validación teórica de las métricas. La validación formal nos ayuda a saber dónde y cómo aplicar las métricas. Hay dos tendencias principales en la validación: los marcos basados en aproximaciones axiomáticas (Weyuker, 1998; Briand et al., 1997; Morasca and Briand, 1997) y los que se basan en la teoría de la medida (Withmire, 1997; Zuse, 1998). La fuerza de la teoría de la medida es la formulación de condiciones empíricas desde las que podemos derivar hipótesis de realidad. La información final que se obtiene al aplicar este tipo de marcos es saber la escala a la que pertenece la métrica y basándonos en esta información poder saber que estadísticos y que transformaciones pueden hacerse con esta métrica.
- Validación empírica.** El objetivo de esta etapa es probar la utilidad práctica de las métricas propuestas. Aunque hay diversas formas de realizar este paso, básicamente, podemos dividir la validación empírica en experimentación y casos de estudio (Basili, Shul y Lanubile, 1999; Fenton y Pfleeger, 1997, Wohlin et al. 2000).

Tal y como se muestra en la figura 2, el proceso de definición y validación de las métricas es evolucionario e iterativo. Como resultado de la retroalimentación, las métricas deben ser redefinidas de acuerdo a las validaciones, teóricas o empíricas, fallidas. En este momento nos encontramos en la primera iteración de definición y validación de métricas de calidad para almacenes de datos.

### 3. Métricas para Almacenes de datos

Teniendo cuenta las características expuestas previamente, hemos realizado un proceso de definición de métricas para almacenes de datos. Estas métricas pueden ser aplicadas a nivel de tabla, estrella o esquema. En este trabajo sólo presentamos las métricas propuestas para esquemas de data warehouse (tabla 1), que son las que hemos validado empíricamente hasta el momento. En (Calero et al., 2001b) se pueden encontrar métricas para los otros niveles.

<b>NFT(Sc).</b> Número de tablas de hechos del esquema.	<b>NDT(Sc).</b> Número de tablas dimensionales del esquema.
<b>NSDT(Sc).</b> Número de tablas dimensionales compartidas por más de una estrella del esquema.	<b>NT(Sc).</b> Número de tablas. Número de tablas de hechos más el número de tablas de dimensión del esquema.
<b>NAFT(Sc).</b> Número de atributos de las tablas de hechos del esquema. $NAFT(Sc) = \sum_{i=1}^{NFT} NA(FT_i)$ Donde FT <sub>i</sub> es la tabla de hechos número de i del esquema Sc	<b>NADT(Sc).</b> Número de atributos de las tablas de dimensión del esquema. $NADT(Sc) = \sum_{i=1}^{NDT} NA(DT_i)$ Donde DT <sub>i</sub> es la tabla dimensional número de i del esquema Sc
<b>NASDT(Sc).</b> Número de atributos de las tablas de dimensión compartidas. $NASDT(Sc) = \sum_{i=1}^{NSDT} NA(DT_i)$ Donde DT <sub>i</sub> es la tabla dimensional número de i del esquema Sc	<b>NA(Sc).</b> Número de atributos del esquema. $NA(Sc) = NAFT(Sc) + NADT(Sc)$
<b>NFK(Sc).</b> Número de claves ajenas en todas las tablas de hechos del esquema. $NFK(Sc) = \sum_{i=1}^{NFT} NFK(FT_i)$ Donde FT <sub>i</sub> es la tabla de hechos número de i del esquema Sc	<b>RSDT(Sc).</b> Ratio de tablas de dimensión compartidas. Es el número de tablas de dimensión, que pertenecen a más de una estrella. $RSDT(Sc) = \frac{NSDT(Sc)}{NDT(Sc)}$
<b>RT(Sc).</b> Ratio de Tablas. Cantidad de tablas de dimensión por tabla de hechos. $RT(Sc) = \frac{NDT(Sc)}{NFT(Sc)}$	<b>RScA(Sc).</b> Ratio de atributos del esquema. Número de atributos en las tablas de dimensión por atributos de las tablas de hechos. $RScA(Sc) = \frac{NADT(Sc)}{NAFT(Sc)}$
<b>RFK(Sc).</b> Ratio de claves ajenas. Cantidad de atributos que son clave ajena. $RFK(Sc) = \frac{NFK(Sc)}{NA(Sc)}$	<b>RSDTA(Sc).</b> Ratio de atributos de las tablas de dimensión compartidas. Cantidad de atributos del esquema que son compartidos. $RSDTA(Sc) = \frac{NASDT(Sc)}{NA(Sc)}$

Tabla 1. Métricas para esquemas de almacenes de datos

### 4. Ejemplo

La figura 3 muestra el ejemplo de un almacén de datos (Adamson y Venerable, 1998). Los valores para las métricas se muestran en la tabla 2.

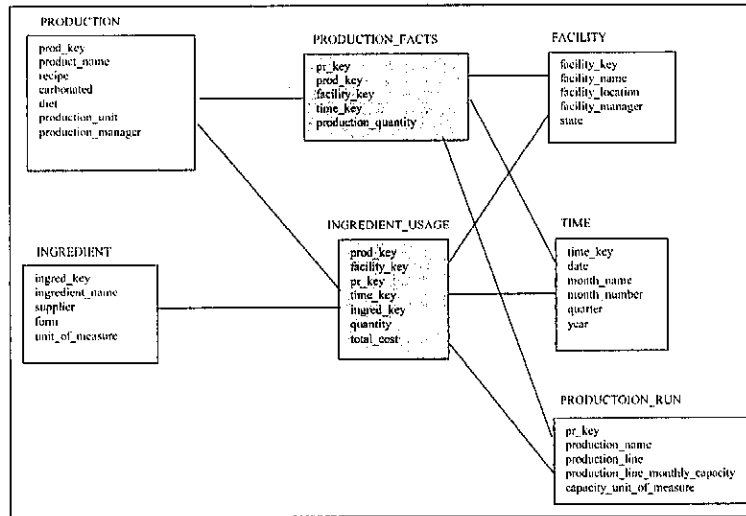


Figura 3. Diseño en estrella de un almacén de datos (Adamson y Venerable, 1998)

Métrica	Valor
NA	40
NFK	9
NDT	5
NT	7
NADT	28
NAFT	12
RFK	9/40

Métrica	Valor
NFT	2
NSDT	4
NASDT	23
RSdT	4/5
RT	5/2
RScA	28/12
RSDTA	23/40

Tabla 2. Valores de las métricas

### 5. Validación formal de las métricas

En esta sección presentaremos la validación formal de las métricas usando el marco formal propuesto por Zuse (1998). Este marco está basado en la teoría de la medida y, por tanto, su objetivo es determinar la escala a la que una métrica pertenece. Sólo mostraremos el proceso completo de formalización en este marco para una de las métricas propuestas. El resto de la validación se realiza de una manera similar y los resultados obtenidos para todas las métricas propuestas se presentarán en la tabla 3. El marco formal de Zuse (1998) trabaja con tres estructuras matemáticas principales, dependiendo cual de ellas cumple una métrica, podremos caracterizarla en una escala. Estas tres estructuras (tabla 3) son: la estructura extensiva, las condiciones de independencia y la relación de creencia modificada. Todos los detalles acerca de estas tres estructuras y el marco formal completo puede ser encontrado en Zuse (1998).

Cuando una medida cumple la estructura extensiva, también cumple las condiciones de independencia y puede ser usada en la escala de ratio.

Si una medida no satisface la estructura extensiva modificada, la regla de combinación (que describe claramente las propiedades de medida del software) existirá o no dependiendo de las condiciones de independencia. Cuando una medida asume las condiciones de independencia pero no la estructura extensiva modificada, el tipo de escala es ordinal.

Cuando una métrica no cumple la estructura extensiva ni tampoco las condiciones de independencia, pero cumple la relación modificada de creencia, puede ser caracterizada como "por encima" del nivel de la escala ordinal (la caracterización de las medidas por encima de la escala ordinal es muy importante porque no podemos hacer prácticamente nada con los números ordinales).

ESTRUCTURA EXTENSIVA MODIFICADA	CONDICIONES DE INDEPENDENCIA	RELACIÓN MODIFICADA DE CREENCIA
<p><b>Axioma1:</b> <math>(A, * \geq)</math> (Orden débil)</p> <p><b>Axioma2:</b> <math>A1 \circ A2 * \geq A1</math> (positividad)</p> <p><b>Axioma3:</b> <math>A1 \circ (A2 \circ A3) = (A1 \circ A2) \circ A3</math> (asociatividad débil)</p> <p><b>Axioma4:</b> <math>A1 \circ A2 = A2 \circ A1</math> (conmutatividad débil)</p> <p><b>Axioma5:</b> <math>A1 * \geq A2 \Rightarrow A1 \circ A * \geq A2 \circ A</math> (monotonicidad débil)</p> <p><b>Axioma6:</b> <math>\exists A3 * \geq A4</math> entonces para cada <math>A1, A2</math>, existe un número natural <math>n</math>, tal que <math>A1 \circ nA3 * \geq A2 \circ nA4</math> (Axioma Arquimedeo)</p>	<p><b>C1:</b> <math>A1 = A2 \Rightarrow A1 \circ A = A2 \circ A</math> y <math>A1 * \geq A2 \Rightarrow A \circ A1 = A \circ A2</math></p> <p><b>C2:</b> <math>A1 = A2 \Rightarrow A1 \circ A = A2 \circ A</math> y <math>A1 * \geq A2 \Rightarrow A \circ A1 = A \circ A2</math></p> <p><b>C3:</b> <math>A1 * \geq A2 \Rightarrow A1 \circ A * \geq A2 \circ A</math>, y <math>A1 * \geq A2 \Rightarrow A \circ A1 * \geq A \circ A2</math></p> <p><b>C4:</b> <math>A1 * \geq A2 \Rightarrow A1 \circ A * \geq A2 \circ A</math>, y <math>A1 * \geq A2 \Rightarrow A \circ A1 * \geq A \circ A2</math></p>	<p><b>MRR1:</b> <math>\forall A, B \in \mathcal{S}: A * \geq B \Rightarrow B * \geq A</math> (completitud)</p> <p><b>MRH2:</b> <math>\forall A, B, C \in \mathcal{S}: A * \geq B</math> y <math>B * \geq C \Rightarrow A * \geq C</math> (transitividad)</p> <p><b>MRB3:</b> <math>\forall A \supset B \Rightarrow A * \geq B</math> (axioma de predominio)</p> <p><b>MRH4:</b> <math>\forall (A \supset B, A \cap C = \emptyset) \Rightarrow (A * \geq B \Rightarrow A \cup C * \geq B \cup C)</math> (monotonicidad parcial)</p> <p><b>MRR5:</b> <math>\forall A \in \mathcal{S}: A * \geq 0</math> (positividad)</p>
<p>Como sabemos la relación binaria <math>* \geq</math> es llamada orden débil si es transitivo y completo:  <math>A1 * \geq A2</math>, and <math>A2 * \geq A3 \Rightarrow A1 * \geq A3</math>  <math>A1 * \geq A2</math> or <math>A2 * \geq A1</math></p>	<p>Donde <math>A1 = A2</math> si y sólo si <math>A1 * \geq A2</math> y <math>A2 * \geq A1</math>, y <math>A1 * \geq A2</math> si y sólo si <math>A1 * \geq A2</math> y no <math>(A2 * \geq A1)</math>.</p>	

Tabla 3. Resumen de las estructuras matemáticas del marco formal de Zuse



**5.1. Validación formal de la métrica NFT**

La medida NFT es una correspondencia :  $NFT: S \rightarrow \mathcal{R}$  tal que para todas las relaciones entre  $S_i$  y  $S_j \in S: S_i \bullet \geq S_j \Leftrightarrow NFT(S_i) \geq NFT(S_j)$ .

Podemos definir la regla de combinación para NFT como:

$$NFT(S_1oS_2) = NFT(S_1) + NFT(S_2)$$

Es decir el número de tablas de hecho de la concatenación de dos esquemas es la suma de las tablas de hecho de cada uno de los esquemas.

Vamos a verificar si la métrica NTF cumple todos los axiomas de la estructura extensiva modificada. NFT cumple el primer axioma (orden débil) ya que si tenemos dos esquemas  $S_1$  y  $S_2$  es obvio que  $NFT(S_1) \geq NTF(S_2)$  o  $NFT(S_1) \leq NFT(S_2)$  (completitud) y además si tenemos 3 esquemas  $S_1, S_2$  y  $S_3$ , la transitividad siempre se cumplirá:  $NFT(S_1) \geq NFT(S_2)$  y  $NFT(S_2) \geq NFT(S_3)$  entonces  $NFT(S_1) \geq NFT(S_3)$ .

NFT también cumple la positividad, ya que el número de tablas de hecho de  $S_1$  o  $S_2$  siempre será mayor que el número de tablas de hecho de  $S_1$ . Esto es debido a que un esquema debe tener al menos una estrella y por tanto debe tener al menos una tabla de hecho.

NFT cumple la asociatividad débil, ya que el número de tablas de hecho del esquema concatenado no depende del orden en que concatenemos los esquemas. Además debe cumplirse pues como el número de tablas de hecho de la concatenación es la suma del número de tablas de los esquemas concatenados y la suma de números enteros cumple la asociatividad. Igualmente NFT cumple la conmutatividad débil pues la suma de números enteros es conmutativa.

NFT cumple la monotonicidad débil el número de tablas de hecho de un esquema  $S_1$  es mayor que el número de tablas de hecho de otro esquema  $S_2$ , es obvio que  $NFT(S_1oS) > NFT(S_2oS)$ .

NFT también cumple el axioma arquimediano ya que si tenemos cuatro esquemas de almacenes de datos  $S_1, S_2, S_3, S_4$  y  $NFT(S_3) > NFT(S_4)$  es fácil deducir que existe algun valor de  $n$  tal que  $NFT(S_1 \circ nS_3) > NFT(S_2 \circ nS_4)$

Viendo que la métrica NFT cumple todo los axiomas de la estructura extensiva modificada, podemos concluir que esta métrica está en la escala de ratio.

Como hemos dicho antes, el resultado de la validación formal del resto de las métricas en el marco formal de Zuse se resume en la tabla 4. Es necesario puntualizar que siguiendo a Zuse (1998), todas las métricas definidas como un porcentaje se pueden caracterizar en la escala absoluta.

	NA	NFK	NDT	NT	NADT	NAFT	REF	NFT	NSDT	NASDT	RSDT	RT	RS <sub>CA</sub>	RS <sub>DTA</sub>
Ax. 1	S	S	S	S	S	S		S	S	S				
Ax. 2	N	S	S	S	S	N		S	N	S				
Ax. 3	S	S	S	S	S	S		S	S	S				
Ax. 4	S	S	S	S	S	S		S	S	S				
Ax. 5	N	N	N	S	N	N		S	N	S				
Ax. 6	N	N	N	S	N	N		S	N	S				
I.C. 1	N	N	N		N	N			N					
I.C. 2	N	N	N		N	N			N					
I.C. 3	N	N	N		N	N			N					
I.C. 4	N	N	N		N	N			N					
MRB1	S	S	S		S	S			S					
MRB2	S	S	S		S	S			S					
MRB3	S	S	S		S	S			S					
MRB4	S	S	S		S	S			S					
MRB5	S	S	S		S	S			S					
Scale	ENC ORD	ENC ORD	ENC ORD	RATIO	ENC ORD	ENC ORD	ABS	RATIO	ENC ORD	RATIO	ABS	ABS	ABS	ABS

Tabla 4. Validación formal de las métricas

Como conclusión de la validación formal hemos obtenido que todas nuestras métricas están en la escala ordinal o en alguna superior. Esto significa que sin métricas software formalmente validas, como remarca Zuse (1998).

**6. Validación empírica de las métricas**

Para poder comprobar si un conjunto de métricas es útil a nivel práctico debemos validarlas empíricamente. Como primera aproximación hemos realizado un experimento en el que elegimos a un grupo de 12 personas, con amplios conocimientos de bases de datos, pero sin experiencia en el campo de los almacenes de datos. A cada uno de los sujetos les proporcionamos un cuadernillo con 11 esquemas de almacenes de datos, de distinta complejidad cada uno y les pedimos que los valoraran, según su opinión, de muy fácil (1) a muy difícil(7).

Una vez teníamos realizado el experimento, realizamos un análisis de correlación usando el estadístico Tau<sub>b</sub> de Kendall, de manera que nuestra hipótesis nula es que no existe relación entre la complejidad estimada por los sujetos y cada uno de los valores de las métricas. La hipótesis alternativa es que si existe relación. En la tabla 5, se muestra el resultado del análisis estadístico.

Como podemos observar en la tabla 5, se puede comprobar que existe relación, al 99% de confiabilidad, entre las métricas NFK, NFT y NT y la variable estimada por los sujetos que realizaron el experimento.

		Complejidad				Complejidad	
Tau_b de Kendall	Coeficiente de correlación	NA	0,448 **	Sig. (bilateral)	NA	0,00	
		NADT	0,365 **		NADT	0,00	
		NAFT	0,518**		NAFT	0,00	
		NASDT	0,520 **		NASDT	0,00	
		NDT	0,571 **		NDT	0,00	
		NFK	0,664 **		NFK	0,00	
		NFT	0,688 **		NFT	0,00	
		NSDT	0,581 **		NSDT	0,00	
		NT	0,669 **		NT	0,00	
		RFK	0,520 **		RFK	0,00	
		RSCA	-0,406 **		RSCA	0,00	
		RSDT	0,515 **		RSDT	0,00	
		RSDTA	0,507 **		RSDTA	0,00	
		RT	-0,478 **		RT	0,00	
** La correlación es significativa al nivel 0,01 (bilateral).							

Tabla 5. Análisis de correlación

A pesar de estos resultados, es necesario seguir realizando experimentos pues este ha sido sólo una primera aproximación. En los siguientes experimentos se propondrá la realización de ejercicios, de manera que las respuestas no sean tan subjetivas y se utilizarán otras técnicas de contraste de hipótesis.

## 7. Conclusiones y trabajo futuro

Si consideramos realmente que la información es el activo principal de las organizaciones, una de las obligaciones principales de los profesionales de las tecnologías de la información debe ser la de asegurar su calidad. La calidad de la información puede ser descompuesta en diferentes tipos de "calidades": calidad de la presentación, calidad del sistema gestor del almacén de datos, calidad de los datos, calidad del de los modelos conceptual, lógico y físico del almacén. Nuestro objetivo centra en éste último punto. Aunque se han propuesto algunas líneas guía para diseñar un "buen" modelo multidimensional, se necesitan indicadores más objetivos.

Estamos elaborando un conjunto de métricas válidas para medir la calidad de los almacenes de datos, las cuales puedan ayudar a los diseñadores a elegir la mejor opción entre varios diseños alternativos.

Hemos presentado y validado varias métricas para medir esquemas de almacenes de datos con diseños en estrella. Sin embargo, esto es únicamente el primer paso en el proceso completo de definición de las métricas. Ahora estamos trabajando en la realización de más experimentos para poder validar empíricamente todas las métricas presentadas, lo cual nos puede permitir descartar y refinar dichas métricas. Debemos definir y validar también otras métricas no orientadas a esquemas. También estamos

trabajando en la realización de una herramienta que permita la obtención automática de los valores de las métricas y la representación de los mismos.

## Referencias

- Adamson, Christopher y Venerable, Michael (1998). *Data Warehouse Design Solutions*. John Wiley and Sons, USA.
- Basili, V.R. y Weiss, D. (1984). A methodology for Collecting Valid Software Engineering Data. *IEEE Transactions on Software Engineering*. SE-10. No. 6. pp.728-738.
- Basili, V.R., Shull, F. y Lanubille, F. (1999). Building Knowledge through families of experiments. *IEEE Transactions on Software Engineering*. July/August. No. 4. pp. 456-473
- Boman M, Bubenko J., Johannesson P. y Wangler B. (1997). *Conceptual Modelling*, Prentice Hall.
- Briand, L.C., Morasca, S. y Basili, V. (1996). Property-based software engineering measurement. *IEEE Transactions on Software Engineering*. 22(1). pp.68-85.
- Bouzeghoub, M, Fabret, F. y Galhardas, H. (2000). Datawarehouse refreshment. Capítulo 4 in *Fundamentals of Data Warehouses*. Ed. Springer.
- Cabbibo, L. y Torlone, R (1998). "A logical approach to multidimensional databases". Sixth International Conference on Extending Database Technology (EDBT'98), Valencia. España. *Lecture Notes in Computer Science 1377*, Springer-Verlag, pp 183-197.
- Calero, C., Piattini, M. y Genero, M (2001a), Metrics for controlling database complexity. Chapter III in *Developing quality complex database systems: practices, techniques and technologies*. Becker (ed), Idea Group Publishing.
- Calero, C., Piattini, M., Pascual, C., Serrano, M.A. (2001b), "Towards Data Warehouse Quality Metrics", Aceptado para publicación en 3rd Workshop on Design and Management of Data Warehouses (DMDW'01).
- Cavero, J., Marcos, E. y Piattini, M. "Metodología para el Diseño de Almacenes de Datos: Etapa de Modelado Conceptual", (2001). 4º Encontro para a Qualidade nas Tecnologias de Informação e Comunicações (QUATIC 2001). Lisboa, Portugal.
- English, L., *Information Quality Improvement: Principles, Methods and Management, Seminar*, 5<sup>th</sup> Ed., Brentwood, TN: Information Impact International, Inc., 1996.
- Fenton, N. y Pffeeger, S., S. L. (1997). *Software Metrics: A Rigorous Approach 2nd. Edition*. London, Chapman & Hall.
- Gardner, S.R. (1998). Building the data warehouse. *Communications of the ACM*, Vol. 41, Nr.9, September. Pp. 52-60.
- Golfarelli, M., Maio, D. y Rizzi, S. (1998) "Conceptual design of data warehouses from E/R schemes". *31<sup>st</sup> Hawaii International Conference on System Sciences*.

- Golfarelle, M. Y Rizzi, S. "Designing The Data Warehouse: Key Steps and Crucial Issues". (1999). *Journal of Computer Science and Information Management*, Vol 2, N. 3.
- Hammergren, T. (1996). *Data Warehousing Building the Corporate Knowledge Base*. International Thomson Computer Press, Milford.
- Harinarayan, V., Rajaraman, A. y Ullman, J. D. (1996). *Implementing Data Cubes Efficiently*. *Proc. of the 1996 ACM SIGMOD International Conference on Management of Data*, Jagadish, H. V. and Mumick, I. S. (eds.), pp. 205-216.
- Hansen, G.W. y Hansen, J.V. (1997). *Diseño y gestión de bases de datos, 2ª Edición*. Prentice Hall.
- Inmon, W.H. (1997). *Building the Data Warehouse*, second edition, John Wiley and Sons, USA.
- Jarke, M., Lenzerini, M., Vassiliou, Y. y Vassiliadis, P. (2000). *Fundamentals of Data Warehouses*, Ed. Springer.
- Kelly, S. (1997). *Data Warehousing in Action*. John Wiley & Sons.
- Kimball, R., Reeves, L., Ross, M. y Thornthwaite, W. (1998). *The Data Warehouse Lifecycle Toolkit*, John Wiley and Sons, USA.
- Labio, W., Quass, D. y Adelberg, B. (1997). *Physical Database Design for Data Warehouses*. *Thirteen International Conference on Data Engineering*, IEEE Computer Society, Birmingham, UK, pp. 277-288
- Lindland, O., Sindre, G. y Solvberg, A. "Understanding Quality in Conceptual Modelling", *IEEE Software*, Marzo, Vol. 11 N° 2, 1994, pp. 42-49.
- Piattini, M., Genero, M., Calero, C., Polo, M. y Ruiz, F. (2000). *Database Quality*. In: *Advanced Database Technology and Design*. Diaz, O. and Piattini, M. (eds.) London. Artech House.
- Weyuker, E.J. (1988). *Evaluating software complexity measures*. *IEEE Transactions on Software Engineering*. 14(9), pp.1357-1365.
- Whitmire, S.A. (1997). *Object Oriented Design Measurement*. Ed. Wiley.
- Zuse, H. (1998). *A Framework of Software Measurement*. Berlin. Walter de Gruyter.

## Diseño de un lenguaje de consulta espacial para documentos XML

J. E. Córcoles<sup>1</sup>, P. González<sup>2</sup>

<sup>1</sup>Sección de Tecnología de la Información.  
Instituto de Desarrollo Regional  
Universidad de Castilla-La Mancha  
ALBACETE

corcoles@idr-ab.uclm.es  
<sup>2</sup>Departamento de Informática  
Escuela Politécnica Superior de Albacete  
Universidad de Castilla-La Mancha  
ALBACETE  
pgonzalez@info-ab.uclm.es

**Abstract:** La reciente aparición de eXtensible Markup Language (XML) como un nuevo estándar para la representación e intercambio en la Web ha originado una considerable atención en el mundo de las bases de datos espaciales. Geographical Markup Language (GML) es una especificación XML para la representación de información espacial que ha sido desarrollado por OGC (OpenGIS Consortium). OGC es un consorcio dedicado al desarrollo de estándares para favorecer la interoperabilidad entre diversos sistemas de geoprocésamiento espacial. El objetivo de GML es permitir el intercambio de información espacial en la Web. Con el mismo propósito, existen lenguajes para otras aplicaciones, como por ejemplo CML (Chemical Markup Language) que permite el intercambio de información relacionada con la química. GML permite tener una base de datos con documentos XML que pueden combinar información espacial o no espacial. En este trabajo se presenta un lenguaje de consulta espacial para documentos XML. El lenguaje de consulta tiene una sintaxis select-from-where basada en SQL (Structured Query Language). El lenguaje incluye un conjunto de operadores espaciales, como (disjoint, touches,...) y un conjunto de operadores alfanuméricos (=, >, <, ...) para los datos no espaciales. Este lenguaje permitirá el desarrollo de aplicaciones que consulten información espacial en la Web, sin ser necesario depender de que la información esté almacenada en un sistema de bases de datos espacial concreto.

### 1. Introducción.

Los documentos basados en XML pueden ser usados para una gran variedad de aplicaciones, en las que se incluyen: e-commerce, comunicaciones business-to-business, etc. Sin embargo, en los últimos meses se ha incorporado con fuerza una