

## Patrocinadores



UNIVERSIDAD DE ZARAGOZA



## Entidades Organizadoras

- Adaspain.
- Asociación de Enseñantes Universitarios de la Informática (AENUU).
- Asociación de Técnicos Informáticos (ATI).
- Asociación Española para la Inteligencia Artificial (AEPiA).
- Asociación para la Interacción Persona-Ordenador (AIPO).
- Asociación para el Desarrollo de la Informática Educativa (ADIE).
- Ayuntamiento de Zaragoza.
- Capítulo Español de la IEEE Computational Intelligence Society.
- Comité Español de Automática (CEA).
- Conferencia de Decanos y Directores de Informática (CODDI) de las Universidades Españolas.
- Departamento de Informática e Ingeniería de Sistemas de la Universidad de Zaragoza.
- European Society for Fuzzy Logic and Technology (EUSFLAT).
- Federación de Asociaciones de Ingenieros en Informática (AI2).
- W3C España (World Wide Web Consortium).
- Programa Nacional de Tecnologías Informáticas - Dirección General de Investigación, Ministerio de Educación y Ciencia.
- Red Española de Metaheurísticas.
- Red Española de Minería de Datos y Aprendizaje.
- Sección Española de la European Association for Computer Graphics (EUROGRAPHICS).
- Sociedad de Arquitectura y Tecnología de Computadores (SARTECO).
- Sociedad de Ingeniería del Software y Tecnologías de Desarrollo del Software (SISTEDES).
- Universidad de Zaragoza.

ISBN: 978-84-9732-595-0

CEDI 2007 XII Jornadas de Ingeniería del Software y Bases de Datos | JISBD'07 |

# CEDI 2007

II CONGRESO ESPAÑOL  
DE INFORMÁTICA  
ZARAGOZA SPAINI

AUDITORIO PALACIO DE CONGRESOS  
11 AL 14 DE SEPTIEMBRE DE 2007

## XII Jornadas de Ingeniería del Software y Bases de Datos

| JISBD'07 |



EDITOR

Xavier Franch



## Patrocinadores



## Entidades Organizadoras

- Adaspain.
- Asociación de Enseñantes Universitarios de la Informática (AENUU).
- Asociación de Técnicos Informáticos (ATI).
- Asociación Española para la Inteligencia Artificial (AEPiA).
- Asociación para la Interacción Persona-Ordenador (AIPO).
- Asociación para el Desarrollo de la Informática Educativa (ADIE).
- Ayuntamiento de Zaragoza.
- Capítulo Español de la IEEE Computational Intelligence Society.
- Comité Español de Automática (CEA).
- Conferencia de Decanos y Directores de Informática (CODDI) de las Universidades Españolas.
- Departamento de Informática e Ingeniería de Sistemas de la Universidad de Zaragoza.
- European Society for Fuzzy Logic and Technology (EUSFLAT).
- Federación de Asociaciones de Ingenieros en Informática (AI2).
- W3C España (World Wide Web Consortium).
- Programa Nacional de Tecnologías Informáticas - Dirección General de Investigación, Ministerio de Educación y Ciencia.
- Red Española de Metaheurísticas.
- Red Española de Minería de Datos y Aprendizaje.
- Sección Española de la European Association for Computer Graphics (EUROGRAPHICS).
- Sociedad de Arquitectura y Tecnología de Computadores (SARTECO).
- Sociedad de Ingeniería del Software y Tecnologías de Desarrollo del Software (SISTEDS).
- Universidad de Zaragoza.

ISBN: 978-84-9732-595-0

CEDI 2007 XII Jornadas de Ingeniería del Software y Bases de Datos | JISBD'07 |

# CEDI 2007

II CONGRESO ESPAÑOL  
DE INFORMÁTICA  
ZARAGOZA SPAINI

AUDITORIO PALACIO DE CONGRESOS  
11 AL 14 DE SEPTIEMBRE DE 2007

## XII Jornadas de Ingeniería del Software y Bases de Datos

| JISBD'07 |



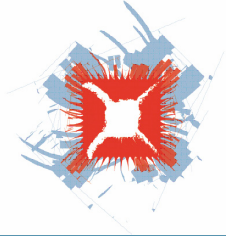
EDITOR

Xavier Franch



**CEDI 2007**  
II CONGRESO ESPAÑOL  
DE INFORMÁTICA  
Nuevos retos  
científicos y tecnológicos  
en Ingeniería Informática  
**ZARAGOZA SPAIN**  
DEL 11 AL 14 DE SEPTIEMBRE

---



# ACTAS DE LAS XII JORNADAS DE INGENIERÍA DEL SOFTWARE Y BASES DE DATOS

**EDITOR**

Xavier Franch

**PATROCINA**

**INTERSYSTEMS**

**COLABORA**

**THOMSON**  
—★—™



**ACTAS DE LAS XII JORNADAS DE INGENIERÍA DEL SOFTWARE Y BASES DE DATOS (JISBD'07)**

No está permitida la reproducción total o parcial de este libro, ni su tratamiento informático, ni la transmisión de ninguna forma o por cualquier otro medio, ya sea electrónico, mecánico, por fotocopia, por registro u otros medios, sin el permiso previo y por escrito de los titulares del Copyright.

Derechos reservados ©2007 respecto a la primera edición en español, por LOS AUTORES  
Derechos reservados ©2007 International Thomson Editores Spain, S.A.

Magallanes, 25; 28015 Madrid, ESPAÑA  
Teléfono 91 4463350  
Fax: 91 4456218  
clientes@parainfo.es

ISBN: 978-84-9732-595-0  
Depósito legal: M-

Maquetación: Los Editores  
Coordinación del proyecto: @LIBROTEX  
Portada: Estudio Dixi  
Impresión y encuadernación: FER Fotocomposición, S. A.

IMPRESO EN ESPAÑA-PRINTED IN SPAIN

## Comité Ejecutivo

### **Presidente del Comité de Programa**

*Xavier Franch (Universitat Politècnica de Catalunya)*

### **Secretario de la Comisión Permanente**

*Mario Piattini (Universidad de Castilla-La Mancha)*

### **Coordinadora de Tutoriales**

*Ana M. Moreno (Universidad Politécnica de Madrid)*

### **Coordinador de Talleres**

*Vicente Pelechano (Universidad Politécnica de Valencia)*

### **Coordinador de Demostraciones**

*Antonio Vallecillo (Universidad de Málaga)*

### **Coordinador de la Sesión de Divulgación de Trabajos Relevantes ya Publicados**

*Oscar Díaz (Universidad del País Vasco)*

### **Composición y Maquetación de Actas**

*Jordi Marco (Universitat Politècnica de Catalunya)*

### **Organización y Relaciones con CEDI 2007**

*Fran J. Ruiz (Universidad de Zaragoza)*

*M. Elena Gómez (Universidad de Zaragoza)*

*Javier Tuya (Universidad de Oviedo)*

## Comité Organizador

### **Presidente del CEDI**

*Alberto Prieto (Universidad de Granada)*

### **Presidente del Comité Científico**

*Juan J. Moreno (Universidad Politécnica de Madrid)*

### **Presidente del Comité Organizador CEDI 2007**

*Victor Viñals (Universidad de Zaragoza)*

### **Coordinador de Actividades Plenarias CEDI 2007**

*José Duato (Universidad Politécnica de Valencia)*

### **Secretario del CEDI 2007**

*José A. Castellanos (Universidad de Zaragoza)*

*José A. Bañares (Universidad de Zaragoza)*

## Comité de Programa

Alberto Abelló, Univ. Polit. Catalunya	Jon Iturrioz, Univ. País Vasco
Silvia Abrahão, Univ. Polit. Valencia	Natalia Juristo, Univ. Polit. Madrid
Jesus Aguilar, Univ. Sevilla	Patricio Letelier, Univ. Polit. Valencia
José Aldana, Univ. Málaga	Antonia Lopes, Univ. Lisboa
Bárbara Álvarez, Univ. Polit. Cartagena	Adolfo Lozano, Univ. Extremadura
María J. Aramburu, Univ. Jaume I	Esperanza Marcos, Univ. Rey Juan Carlos
João Araújo, Univ. Nova de Lisboa	Eduardo Mena, Univ. Zaragoza
Orlando Belo, Univ. do Minho	Ana Moreira, Univ. Nova de Lisboa
Rafael Berlanga, Univ. Jaume I	Juan J. Moreno, Univ. Polit. Madrid
Pere Botella, Univ. Polit. Catalunya	Juan M. Murillo, Univ. Extremadura
Nieves Brisaboa, Univ. Coruña	Oscar Pastor, Univ. Polit. Valencia
Isabel S. Brito, Inst. Polit. Beja	Antonio Polo, Univ. Extremadura
Coral Calero, Univ. Castilla-La Mancha	Carme Quer, Univ. Polit. Catalunya
Carlos Canal, Univ. Málaga	Celia Ramos, Univ. Algarve
José M. Caveró, Univ. Rey Juan Carlos	Isidro Ramos, Univ. Polit. Valencia
Matilde Celma, Univ. Polit. Valencia	José Riquelme, Univ. Sevilla
Rafael Corchuelo, Univ. Sevilla	Antonio Rito, Univ. Técnica de Lisboa
Dolors Costal, Univ. Polit. Catalunya	Antonio Ruíz, Univ. Sevilla
Yania Crespo, Univ. Valladolid	Francisco Ruíz, Univ. Castilla-La Mancha
Oscar Dieste, Univ. Polit. Madrid	José Samos, Univ. Granada
Javier Dolado, Univ. País Vasco	Fernando Sánchez, Univ. Extremadura
João Falcão e Cunha, Univ. Porto	Juan Sánchez, Univ. Polit. Valencia
Pablo de la Fuente, Univ. Valladolid	Ernest Teniente, Univ. Polit. Catalunya
Lidia Fuentes, Univ. Málaga	Miguel Toro, Univ. Sevilla
Mario Gaspar da Silva, Univ. Lisboa	Ambrosio Toval, Univ. Murcia
Marcela Genero, Univ. Castilla-La Mancha	Juan C. Trujillo, Univ. Alicante
Cristina Gómez, Univ. Polit. Catalunya	Javier Tuya, Univ. Oviedo
Jaime Gómez, Univ. Alicante	Belén Vela, Univ. Rey Juan Carlos
Alfredo Goñi, Univ. País Vasco	Cristina Vicente, Univ. Polit. Cartagena
Juan Hernández, Univ. Extremadura	

## Comité Asesor para la Selección de Trabajos de Prestigio

Oscar Díaz (Presidente), Univ. País Vasco	Neil A.M. Maiden, City Univ. London
Alan Davis, Univ. of Colorado	Timos Sellis, Nat. Technical Univ. Athens

## Revisores Adicionales

César J. Acuña  
Amaia Aguirregoitia  
Diego Alonso  
David Benavides  
Jordi Cabot  
Paloma Cáceres  
Javier Cámara  
Dante Carrizo  
Pedro J. Clemente  
Jose M. Conejero  
Javier Cubo  
Norberto Díaz  
Amador Durán  
Sergio España  
Mauricio Espinoza  
Ismael Etxeberria  
Antonio Fariña  
Raul Fernandez  
L. Fredlund  
Antonielly Garcia  
Antonio Cesar Gómez  
Ángel Herranz  
Sergio Ilarri  
Miguel Ángel Laguna  
Maria Lencastre  
Marta López  
Francisco Javier Lucas  
María Esperanza Manso  
Julio Mariño  
José Manuel Marqués  
Francisco Martínez  
Jorge Martínez

Miguel Ángel Martínez  
Fernando Molina  
Ana M. Moreno  
Elena Navarro  
Ismael Navas  
Isabel Nepomuceno  
Juan A. Nepomuceno  
Joaquín Nicolás  
Guadalupe Ortiz  
Juan Angel Pastor  
Joaquin Peña  
Jenifer Pérez  
Juan Manuel Pérez  
Beatriz Pontes  
Álvaro Prieto  
Antonia M. Reina  
Domingo Savio Rodríguez  
Roberto Rodríguez  
Oscar Romero  
Fran J. Ruiz  
Angeles Saavedra  
Gwen Salaün  
Pedro Sánchez  
André L. Santos  
Diego Seco  
Jesús Serrano  
Encarna Sosa  
Toufik Taibi  
Raquel Trillo  
José Antonio Troyano  
Juan Manuel Vara

## **Sistema Automático de Revisión**

*Quercus Software Engineering Group*

Jose Javier Berrocal Universidad de Extremadura

Conferencia auspiciada por





## Prólogo

Respondiendo a su cita anual, las XII Jornadas de Ingeniería del Software y Bases de Datos (JISBD) se han celebrado en Zaragoza, entre el 11 y el 14 de septiembre de 2007. Las Jornadas representan un punto de encuentro de la comunidad investigadora en ingeniería del software y en bases de datos. En sus inicios se celebraron dos eventos diferenciados, las Jornadas de Ingeniería del Software y las Jornadas sobre Investigación y Docencia en Bases de Datos. Posteriormente, en 1999, ambos eventos se unificaron en uno solo, reflejando la interrelación existente entre estas disciplinas. En esta duodécima edición, las Jornadas han constituido, una vez más, un punto de encuentro en el que profesionales y académicos de España, Portugal y Latinoamérica, de ambos campos, han podido compartir experiencias y resultados entre distintos grupos de investigación, desarrollo e innovación tecnológica.

Actualmente, JISBD es un evento auspiciado por Sociedad de Ingeniería del Software y Tecnologías de Desarrollo de Software (SISTEDES, <http://www.sistedes.org>). Entre los fines de dicha organización destacan el de promover la investigación, la innovación y la transferencia de tecnología entre los distintos agentes involucrados en el avance las tecnologías del Software y el de fomentar actividades con otras asociaciones nacionales e internacionales con fines similares, consiguiendo así proporcionar una mayor visibilidad a la investigación de sus asociados.

Al igual que en 2005, las XII Jornadas de Ingeniería del Software y Bases de Datos se han realizado en el marco del II Congreso Español de Informática (CEDI 2007). Esto ha permitido a los participantes de las Jornadas participar en las diversas actividades de CEDI de interés para toda la comunidad de investigación en Informática, tales como conferencias invitadas y mesas redondas. La celebración cada dos años de JISBD en el marco de CEDI encaja con los objetivos citados de dicha organización.

Este volumen recoge los trabajos seleccionados por el Comité de Programa de JISBD'07. Se recibieron un total de 87 contribuciones de 9 países: España, Portugal, Argentina, Brasil, Chile, Colombia, Cuba, México y Venezuela. Cada contribución fue revisada por tres miembros del Comité de Programa. Posteriormente, se abrió una fase de discusión en la que se debatieron en mayor profundidad algunos trabajos y eventualmente se pidieron revisiones adicionales para ellos; asimismo, algunos trabajos se aceptaron condicionalmente, pendientes de verificar que la versión definitiva trataba adecuadamente los comentarios de los revisores; gracias al esfuerzo de los autores, todos estos trabajos fueron finalmente aceptados. Como resultado de todo el proceso, se configuró un programa compuesto por 30 artículos. Adicionalmente, se seleccionaron 5 trabajos más para su presentación como artículos cortos. Además, en esta edición de JISBD se recogió la posibilidad de presentar trabajos ya publicados en foros de prestigio reconocido. Se seleccionaron 4 artículos de esta modalidad. Finalmente, destacamos la celebración de una sesión para la presentación de herramientas, cuya convocatoria tuvo una acogida excelente por parte de la comunidad de JISBD, de manera que en dicha sesión se programaron un total de 19 demostraciones de herramientas.

El día previo a la conferencia, se organizaron un total de 7 talleres y un tutorial. Estos eventos están ganando importancia a cada nueva edición de JISBD y en el caso de los talleres, están creando sus propias comunidades con intereses más específicos. Algunos talleres ya están plenamente consolidados y llegan a acumular hasta un total de 8 ediciones. Cabe destacar que a partir de este año, las actas de los talleres se recogen en una publicación única en formato electrónico, con el soporte de SISTEDES, para potenciar la difusión de los trabajos presentados.

En referencia al programa, mencionar la participación de dos conferenciantes invitados de reconocido prestigio, siguiendo la pauta de ediciones anteriores. La primera conferencia impartida por Stephen Mellor, miembro del Object Management Group, y con un largo historial en la formulación de métodos para el análisis orientado a objetos. La segunda conferencia a cargo del profesor

John Mylopoulos, que posee igualmente una dilatada experiencia en diversos ámbitos de la ingeniería del software. La presencia de estos dos investigadores representó un elemento importante en el programa de las Jornadas.

Quisiera destacar un hecho que no por obvio, deja de ser merecedor de mención. La celebración de un evento de las características de JISBD, con una participación cada vez más numerosa y consolidada, y con unas exigencias de calidad que se van incrementando en cada edición, no podría realizarse sin la dedicación totalmente desinteresada de un gran número de personas. Desde el punto de vista científico, el trabajo en equipo desarrollado por los miembros del Comité Ejecutivo, en cuyo seno se han debatido los temas más candentes en la configuración de la oferta científica del congreso; y por supuesto la ardua y puntual labor de revisión efectuada por los miembros del Comité de Programa y los revisores adicionales. Desde el punto de vista organizativo, destacar la gran dedicación de los miembros del Comité Ejecutivo responsables de las tareas de enlace con CEDI, y la labor del Grupo Quercus de Ingeniería del Software de la Universidad de Extremadura, quienes han estado a cargo de todo el sistema de recepción y revisión de artículos. También deseo agradecer el soporte recibido por las entidades patrocinadoras y colaboradoras, y en especial la labor de respaldo de SISTEDES, tanto por lo que se refiere a apoyo logístico como a tareas de difusión, como ya se ha comentado. Y por último, especialmente, a los autores de los trabajos enviados a JISBD'07, en definitiva son ellos los que hacen posible la celebración del evento.

Finalmente, desear que el volumen que ahora tienes en tus manos, y que refleja el estado del arte en la investigación en Ingeniería del Software y Bases de Datos en la comunidad de habla hispana y portuguesa, sea de utilidad para tu trabajo.

Zaragoza, Septiembre 2007  
Xavier Franch (editor)



<b>Índice</b>	<b>9</b>
---------------	----------

## Índice

### CONFERENCIAS INVITADAS

<b>Creativity, Automation and Technology</b>	
<i>Stephen J Mellor</i> . . . . .	15
<b>Goal-Oriented Requirements Engineering</b>	
<i>John Mylopoulos</i> . . . . .	17

### TUTORIAL

<b>Tutorial: Herramientas Eclipse para Desarrollo de Software Dirigido por Modelos</b>	
<i>Cristina Vicente-Chicote y Diego Alonso</i> . . . . .	21

### TRABAJOS RELEVANTES YA PUBLICADOS

<b>Access Control and Audit Model for the Multidimensional Modeling of Data Warehouses</b>	
<i>Eduardo Fernández-Medina, Juan Trujillo, Rodolfo Villarroel y Mario Piattini</i> . . . . .	25
<b>A UML profile for multidimensional modeling in data warehouses</b>	
<i>Sergio Luján-Mora, Juan Trujillo e Il-Yeol Song</i> . . . . .	26
<b>Location-Dependent Queries in Mobile Contexts: Distributed Processing Using Mobile Agents</b>	
<i>Sergio Ilarri, Eduardo Mena y Arantza Illarramendi</i> . . . . .	27
<b>Integrating techniques and tools for testing automation</b>	
<i>Macario Polo, Sergio Tendero y Mario Piattini</i> . . . . .	28

### DESARROLLO DE SOFTWARE DIRIGIDO POR MODELOS

<b>Utilidad de las transformaciones modelo-modelo en la generación automática de código</b>	
<i>Javier Luis Cánovas Izquierdo, Óscar Sánchez Ramón, Jesús Sánchez Cuadrado y Jesús García Molina</i> . . . . .	31
<b>Building Ubiquitous Business Process following an MDD approach</b>	
<i>Pau Giner, Victoria Torres y Vicente Pelechano</i> . . . . .	41
<b>A case study on modeling persistence with MDA tools</b>	
<i>Giuliano Luz Pigatti Caliarì y Paulo Sérgio Muniz Silva</i> . . . . .	51

### ALMACENES Y MINERÍA DE DATOS

<b>Ingeniería inversa dirigida por modelos para el diseño de almacenes de datos</b>	
<i>Jose-Norberto Mazón, Enrique Ortega y Juan Trujillo</i>	63
<b>Minería de datos con clustering en espacios multidimensionales mediante modelos conceptuales extendiendo UML</b>	
<i>Jose Zubcoff, Jesús Pardillo y Juan Trujillo</i>	73
<b>Una extensión del metamodelo relacional de CWM para representar Almacenes de Datos Seguros a nivel lógico</b>	
<i>Emilio Soler, Juan Trujillo, Eduardo Fernández-Medina y Mario Piattini</i>	83

### PRUEBAS DEL SOFTWARE

<b>Generación sistemática de pruebas para composiciones de servicios utilizando criterios de suficiencia basados en transiciones</b>	
<i>José García-Fanjul, Javier Tuya y Claudio de la Riva</i>	95
<b>Generación automática de objetivos de prueba a partir de casos de uso mediante partición de categorías y variables operacionales</b>	
<i>Javier J. Gutiérrez, María J. Escalona, Manuel Mejías, Jesús Torres y Arturo Torres-Zenteno</i>	105
<b>370.000 bugs del proyecto Debian pueden ser analizados usando btsextract</b>	
<i>Miguel Pérez Francisco y Pablo Boronat Pérez</i>	115

### TECNOLOGÍAS DE BASES DE DATOS

<b>Búsqueda de vecinos en espacios multidimensionales agujereados</b>	
<i>Manuel Barrena, Carlos Pachón y Elena Jurado</i>	125
<b>Indexación dinámica para la recuperación de información basada en búsqueda por similitud</b>	
<i>Nieves R. Brisaboa, Antonio Fariña, Oscar Pedreira y Nora Reyes</i>	134
<b>WCSA: Un autoíndice orientado a palabras para textos en lenguaje natural</b>	
<i>Eduardo Rodríguez, Antonio Fariña, Ángeles S. Places, José R. Paramá y Oscar Pedreira</i>	144

### LÍNEAS DE PRODUCTO. ORIENTACIÓN A ASPECTOS

<b>Variabilidad, Trazabilidad y Líneas de Productos: una Propuesta basada en UML y Clases Parciales</b>	
<i>Miguel A. Laguna y Bruno González-Baixauli</i>	157
<b>Verificación de Modelos Arquitectónicos Orientados a Aspectos</b>	
<i>Jennifer Pérez, Cristóbal Costa, Jose Ángel Carsí e Isidro Ramos</i>	167
<b>Gestión Integral de Requisitos de Seguridad en Líneas de Producto Software</b>	
<i>Daniel Mellado, Eduardo Fernández-Medina y Mario Piattini</i>	177



## REQUISITOS. METAMODELADO EN MEDICIÓN

<b>Una metodología para elicitación de requisitos en proyectos GSD</b> <i>Gabriela N. Aranda, Aurora Vizcaíno, Alejandra Cechich, Mario Piattini y Juan Pablo Soto</i>	191
<b>Una Aproximación de Metamodelado para la Evaluación de Calidad en Procesos de Desarrollo Web</b> <i>Cristina Cachero, Emilio Insfran, Silvia Abrahão y Geert Poels</i>	201
<b>Marco de Trabajo basado en MDA para la Medición Genérica del Software</b> <i>Beatriz Mora, Félix García, Francisco Ruiz, Mario Piattini, Artur Boronat, Abel Gómez, José Á. Carsí e Isidro Ramos</i>	211

## MODELIZACIÓN CONCEPTUAL DE DATOS

<b>Definición, importancia y especificación en UML de las restricciones de integridad constante y permanente</b> <i>Raquel Pau y Antoni Olivé</i>	223
<b>Modelado de Aplicaciones Web Reactivas al Usuario</b> <i>Irene Garrigós y Jaime Gómez</i>	232
<b>Towards Integration of Access Control in the Hypermedia Development Process</b> <i>Daniel Sanz, Paloma Díaz e Ignacio Aedo</i>	242

## ARQUITECTURAS SOFTWARE

<b>Diseño de Sistemas Groupware sobre una Arquitectura centrada en Servicios Cooperativos: Ágora</b> <i>Miguel A. Martínez-Prieto, Pablo de la Fuente y Carlos E. Cuesta</i>	255
<b>Una Propuesta de Libro Electrónico basada en Composición de Responsabilidades sobre la Estructura Lógica</b> <i>Miguel A. Martínez-Prieto, Pablo de la Fuente, Jesús Vegas y Joaquín Adiego</i>	265
<b>Recuperación y procesado de datos biológicos mediante Ingeniería Dirigida por Modelos</b> <i>Abel Gómez, Artur Boronat, Claudia Täubner, Jose Á. Carsí, Isidro Ramos y Silke Eckstein</i>	275

## MODELOS DE CALIDAD

<b>Evaluando la Calidad de los Datos en Portales Web</b> <i>Angélica Caro, Coral Calero y Mario Piattini</i>	287
<b>Una propuesta de un modelo conceptual de calidad de almacenes de datos</b> <i>Manuel Serrano, Rafael Romero, Jose-Norberto Mazón, Juan Trujillo y Mario Piattini</i>	297
<b>Evaluación de los niveles de calidad en las transformaciones de modelos basado en el estudio de factores de éxito</b> <i>Alejandro Gómez, Gustavo Muñoz y Juan Carlos Granja</i>	307

## PROCESOS

<b>Técnica de Mejora del Mantenimiento Software Basada en Valor</b> <i>Daniel Cabrero, Javier Garzás y Mario Piattini</i> . . . . .	317
<b>Modelo para la Implementación de Mejora de Procesos en Pequeñas Organizaciones Software</b> <i>Francisco J. Pino, Juan C. Vidal, Félix Garcia y Mario Piattini</i> . . . . .	326
<b>Especificación de Procesos de Negocio Seguros a través de una extensión de UML 2.0</b> <i>Alfonso Rodríguez, Eduardo Fernández-Medina, Mario Piattini y Juan Trujillo</i> . . . . .	336

## ARTÍCULOS CORTOS

<b>Eficacia del método ELVIRA - Relato de un experimento</b> <i>Montse Ereño y Rebeca Cortazar</i> . . . . .	349
<b>Tracking the Evolution of Feature Oriented Product Lines</b> <i>Salvador Trujillo, Gentzane Aldekoa y Goiuri Sagardui</i> . . . . .	355
<b>Transformaciones QVT para la obtención de Clases de Análisis a partir de un Modelo de Proceso de Negocio Seguro</b> <i>Alfonso Rodríguez, Ignacio García, Eduardo Fernández-Medina y Mario Piattini</i> . . . . .	361
<b>Definición de un Proceso para la Construcción de Refactorizaciones</b> <i>Raúl Marticorena, Carlos López y Yania Crespo</i> . . . . .	367
<b>Combinando Modelos de Procesos y Activos Reutilizables en una Transición poco Invasiva hacia las Líneas de Producto de Software</b> <i>Orlando Avila-García, Antonio Estévez García, E. Victor Sánchez Rebull y José Luis Roda García</i> . . . . .	373

## DEMOSTRACIONES

<b>Generation of Business Process based Web Applications</b> <i>Pau Giner, Victoria Torres y Vicente Pelechano</i> . . . . .	381
<b>PervGT: Herramienta CASE para la Generación Automática de Sistemas Pervasivos</b> <i>Estefanía Serral, Carlos Cetina, Javier Muñoz y Vicente Pelechano</i> . . . . .	383
<b>UMLtoCSP: Una herramienta para la verificación de modelos UML/OCL mediante Constraint Programming</b> <i>Jordi Cabot, Robert Clarisó, Patricia de la Fuente Y Daniel Riera</i> . . . . .	385
<b>MDBE: Una Herramienta Automática para el Modelado Multidimensional</b> <i>Oscar Romero y Alberto Abelló</i> . . . . .	387
<b>MOMENT CASE: Un prototipo de herramienta CASE</b> <i>Abel Gómez, Artur Boronat, Jose Á. Carsí e Isidro Ramos</i> . . . . .	389
<b>Comprobación eficiente de restricciones de integridad en OCL</b> <i>Jordi Cabot y Ernest Teniente</i> . . . . .	391
<b>The MOVA Tool: A Rewriting-Based UML Modeling, Measuring, and Validation Tool</b> <i>Manuel Clavel, Marina Egea y Viviane Torres da Silva</i> . . . . .	393



<b>Demostración de la herramienta AGE (Agile Generative Environment)</b>	
<i>Jesús Sánchez Cuadrado y Jesús García Molina . . . . .</i>	395
<b>ModelSET: Soporte a Edición y Transformaciones de Modelos</b>	
<i>Antonio Estévez García, E. Victor Sánchez Rebull, Francisco Vargas Ruiz, Orlando Avila-García, Adolfo Sánchez-Barbudo Herrera y José Luis Roda García . . . . .</i>	397
<b>PRISMA CASE</b>	
<i>Jennifer Pérez, Cristóbal Costa, Jose A. Carsí e Isidro Ramos . . . . .</i>	399
<b>StateML: modelado gráfico de máquinas de estados y generación de código siguiendo un enfoque MDE</b>	
<i>Cristina Vicente-Chicote, Diego Alonso y Bárbara Álvarez . . . . .</i>	401
<b>V<sup>3</sup> Studio: Un entorno gráfico para el diseño de sistemas basados en componentes siguiendo un enfoque dirigido por modelos</b>	
<i>Cristina Vicente-Chicote, Diego Alonso y Olivier Barais . . . . .</i>	403
<b>REMM-Studio: Un entorno integrado para dar soporte a un enfoque de Ingeniería de Requisitos Dirigido por Modelos</b>	
<i>Cristina Vicente-Chicote, Begoña Moros y Ambrosio Toval . . . . .</i>	405
<b>MORPHEUS: support from AO-Requirements to AO-Software Architecture</b>	
<i>Elena Navarro, Patricio Letelier e Isidro Ramos . . . . .</i>	407
<b>Maudeling: Herramienta de gestión de modelos usando Maude</b>	
<i>José E. Rivera, Francisco Durán, Antonio Vallecillo y J. Raúl Romero . . . . .</i>	409
<b>WebTE: Generación de aplicaciones Web dirigida por modelos</b>	
<i>Santiago Meliá , Jaime Gómez y Jose Luis Serrano . . . . .</i>	411
<b>CE4WEB: Una Herramienta CASE Colaborativa para el Modelado de Aplicaciones con UML</b>	
<i>Víctor M.R. Penichet, María D. Lozano, J.A. Gallud y R. Tesoriero . . . . .</i>	413
<b>MaCMAS CASE Tool Demonstration: MDD-based refinement of Collaboration-Based UML Models</b>	
<i>Joaquín Peña y Antonio Ruiz-Cortés . . . . .</i>	415
<b>FAMA:hacia el análisis automático de modelos de características</b>	
<i>Pablo Trinidad, David Benavides, Sergio Segura y Antonio Ruiz Cortés . . . . .</i>	417

Angélica Caro

Departamento de Ciencias de la Computación  
y Tecnologías de Información,  
Universidad del Bío Bío  
Chillán, Chile  
mcaro@ubiobio.cl

Coral Calero y Mario Piattini

Grupo de Investigación Alarcos,  
Departamento de Tecnologías y Sistemas de Información,  
Instituto de Investigación y Desarrollo UCLM-SOLUZIONA  
Universidad de Castilla-La Mancha, Ciudad Real, España  
{Coral.Calero, Mario.Piattini}@uclm.es

## Resumen

Los avances tecnológicos y el uso de Internet han favorecido la aparición de una gran variedad de aplicaciones Web, entre ellas los Portales Web. A través de éstos, muchas personas obtienen los datos necesarios para desarrollar sus tareas y/o tomar decisiones necesitando, por tanto, estar seguros de que los datos que ofrecen estas aplicaciones tienen el nivel de calidad apropiado para el uso que desean darles.

En este artículo se presenta el desarrollo de un Modelo de Calidad de Datos para Portales Web (PDQM), centrado en la perspectiva del usuario. PDQM fue desarrollado en dos fases, la primera de ellas generó un modelo teórico en el cual se identificaron 33 atributos de calidad de datos aplicables en un portal Web. La segunda, donde este modelo teórico es transformado en uno operacional. Es resultado es, un modelo que realmente puede ser usado en un proceso de evaluación de la calidad de los datos de un portal Web.

## 1. Introducción

Un portal Web (PW) es un sitio que agrega información desde múltiples fuentes de la Web y que la organiza de manera que se facilite su uso [24]. Un PW, en general, provee a los usuarios (o consumidores de datos) por medio de una interfaz simple, acceso a diferentes fuentes [15], acceso a la información en línea y servicios relacionados con la información [25]. Estos últimos años, se ha incrementado el número de organizaciones que han recurrido a PW para complementar, sustituir o extender sus servicios a los clientes [25]. Asimismo, cada día aumenta el número de usuarios que usan PW en busca de datos para desarrollar sus tareas y/o tomar decisiones. Estos usuarios

necesitan asegurarse que los datos obtenidos son apropiados para el uso que desean darles. Es decir, necesitan saber si los datos tienen un nivel apropiado de calidad.

El concepto de Calidad de Datos o Información (en este artículo los términos Dato e Información serán usados como sinónimos) a menudo es definido como datos que son "apropiados para el uso" [22], es decir, la capacidad de una colección de datos de reunir los requisitos de los usuarios [3, 22]. Esto hace que sea un concepto relativo, ya que lo que es apropiado para unos usuarios podría no serlo para otros [23].

En los últimos años, los investigadores del área de calidad de datos (DQ en siglas inglesas) han comenzado a interesarse por abordar la DQ en la Web [9]. Sin embargo, a pesar que algunos trabajos reconocen la DQ como un factor relevante en la calidad de un PW [25], hasta ahora no hay propuestas que se concentren en la DQ en PW.

En este artículo, presentamos un Modelo de Calidad de Datos para PW (PDQM, Portal Data Quality Model), centrado en la perspectiva del consumidor de datos. Este modelo ha sido desarrollado en dos partes. La primera, correspondiente a su definición teórica, dio como resultado un conjunto de 33 atributos de DQ aplicables en un PW. En la segunda, el modelo teórico se transforma en uno operacional a través de la definición de una estructura donde organizar los atributos de DQ y las medidas necesarias para aplicar el modelo en un proceso de evaluación de DQ. Considerando la doble naturaleza de la DQ (objetiva y subjetiva), para realizar esta conversión, hemos utilizado un enfoque probabilístico basado en redes bayesianas (BN en siglas inglesas) [16] con la idea de obtener un modelo flexible, adaptable al contexto de evaluación y a las preferencias de los consumidores de datos.

El resto del artículo se organiza de la siguiente



forma: la sección 2 describe el método usado para construir PDQM, la sección 3 describe el desarrollo del modelo teórico, la sección 4 muestra la conversión de PDQM en un modelo operacional y, finalmente, la sección 5 presenta nuestras conclusiones y trabajo futuro.

## 2. Método de desarrollo de PDQM

Como muestra la Figura 1, el desarrollo de PDQM fue organizado en dos partes:

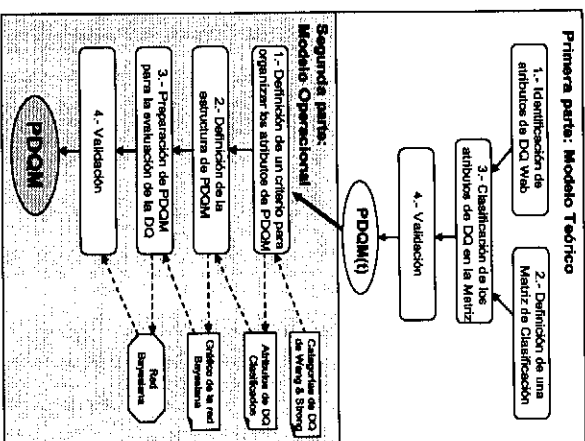


Figura 1. Proceso de desarrollo de PDQM

La primera parte de este proceso, que se divide en cuatro etapas, consistió en la definición teórica del modelo y reúne los aspectos claves que representan la perspectiva del consumidor de datos y las principales características de los PW.

Partiendo del modelo teórico, compuesto por 33 atributos de DQ, el objetivo de la segunda parte fue convertir este en un modelo operacional.

A continuación las secciones 3 y 4 explican en forma detallada el desarrollo del modelo teórico y su transformación a un modelo operacional.

## 3. Desarrollo del modelo Teórico

Para la definición teórica de PDQM se consideraron tres aspectos claves:

- *La perspectiva del consumidor de datos.* Representada en nuestro modelo por las expectativas de DQ de los consumidores de datos en Internet, propuestas en [21]. Estas expectativas están organizadas en 6 categorías: Privacidad, Contenido, Calidad de los valores, Presentación, Mejora y Compromiso.
- *Atributos de DQ Web.* Asumiendo que los PW son aplicaciones web, se consideró útil aprovechar los trabajos previos desarrollados en distintos dominios del contexto Web, y en los cuales se propusieron atributos de DQ.
- *Funcionalidades básicas de un portal Web.*

Para la definición de PDQM, hemos considerado que los consumidores de datos juzgan la DQ a través del uso de las funciones de software básicas que ofrece un PW. Y hemos usado como funcionalidades básicas las propuestas por Collins en [5]. Estas funcionalidades son las siguientes: Integración y Puntos de Datos, Taxonomía, Capacidades de Búsqueda, Características de Ayuda, Gestión de Contenido, Procesos y Acciones, Colaboración y Comunicación, Personalización, Presentación, Administración y Seguridad. Estos tres aspectos son la base para el desarrollo el modelo teórico. Las siguientes subsecciones, describen el desarrollo de éste.

### 3.1. Identificación de atributos de DQ Web

Considerando que no existían trabajos que abordaran específicamente la DQ en los PW, pero asumiendo que si los había en el contexto de la Web, la primera etapa consistió en estudiar la literatura del área de DQ en la Web y obtener de ella atributos de DQ. Para hacer esto, desarrollamos una revisión sistemática de la literatura del área [13]. Esta revisión abarcó un total de 55 trabajos (entre 1995 y 2004), entre los cuales seleccionamos 12 que abordan la DQ en diversos dominios de la Web (sitios Web [7, 11, 18], integración de datos [2, 19], comercio electrónico [12], portales Web [25], servicios electrónicos cooperativos [8], toma de decisiones [10], redes organizacionales [17], DQ en la Web [9]).

Como resultado del análisis de estas propuestas obtuvimos 100 atributos de DQ Web. Con la idea de reducir este número, y habiendo detectado atributos repetidos y sinónimos unificamos aquellos que coincidían en nombre y/o significado.

Funcionalidad	Accesabilidad	Actualidad	Aplicabilidad	Atractivo	Cantidad de datos	Complejidad	Confiable	Consistencia interna	Costo efectivo	Credibilidad	Disponibilidad	Documentación	Duplicidad	Entendibilidad	Especialización	Exactitud	Expiración	Facilidad de operación	Flexibilidad	Granularidad	Información de la fuente	Interactividad	Interpretabilidad	Latencia	Mantenibilidad	Novedad	Objetividad	Ontología	Oportunidad	Organización	Precio	Relevancia	Representación concisa	Representación consistente	Reputación	Seguridad	Soporte de usuario	Tiempo de respuesta	Trazabilidad	Validez	Valor agregado	Total de atributos			
Integración y puntos de Datos	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	16			
Taxonomía	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11		
Capacidades de Búsqueda	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	15		
Características de Ayuda	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	8		
Procesos y Acciones	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	28		
Comunicación y Colaboración	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	6		
Personalización	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	7		
Administración	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	18		
Seguridad	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	6		
Momento de Revisión	7	6	2	1	9	5	7	0	0	6	3	1	1	1	1	3	4	1	8	1	1	0	1	2	5	0	0	3	2	1	0	5	4	0	7	9	2	2	5	8	2	7	8	1	1

Tabla 1. Atributos de Calidad de Datos de PDQM.

Con esto, obtuvimos un conjunto final de 41 atributos de DQ (ver fila superior en la Tabla 1).

### 3.2. Definición de una matriz de clasificación

Junto con la identificación de los atributos de DQ Web, definimos una matriz que nos permitiera hacer un análisis preliminar de la aplicabilidad de estos atributos en un PW. La matriz construida refleja la relación existente entre las funcionalidades de un PW identificadas en [5] (columnas de la matriz) y las expectativas de DQ de los consumidores de datos en Internet propuestas en [21] (filas de la matriz). Sobre esta matriz analizamos, por cada funcionalidad, las expectativas de DQ aplicables. La Figura 2, representa la matriz y destaca mediante el símbolo "✓", las relaciones establecidas entre funcionalidad y expectativa.

Funcionalidades de un Portal Web	Privacidad	Comentarios	Calidad de los valores	Presentación	Independencia	Compartido
Integración y puntos de datos	✓	✓	✓	✓	✓	✓
Taxonomía	✓	✓	✓	✓	✓	✓
Capacidades de búsqueda	✓	✓	✓	✓	✓	✓
Características de Ayuda	✓	✓	✓	✓	✓	✓
Gestión de Contenido	✓	✓	✓	✓	✓	✓
Procesos y Acciones	✓	✓	✓	✓	✓	✓
Comunicación y Colaboración	✓	✓	✓	✓	✓	✓
Personalización	✓	✓	✓	✓	✓	✓
Administración	✓	✓	✓	✓	✓	✓
Seguridad	✓	✓	✓	✓	✓	✓
Categorías de Expectativas de Datos de los Consumidores de Datos	✓	✓	✓	✓	✓	✓

Figura 2. Matriz de clasificación de atributos de DQ

### 3.3. Clasificación de atributos de DQ

Esta etapa consistió en analizar la aplicabilidad de cada uno de los 41 atributos de DQ Web, obtenidos en la etapa 1, en cada relación (funcionalidad, expectativa) establecida en la matriz de clasificación. Esto se hizo asumiendo el rol del consumidor de datos y considerando la definición de cada atributo. Como resultado de esto, se determinó que 34 atributos de DQ eran aplicables en PW. En la Tabla 1 se muestra con un color más oscuro un resumen de los atributos de DQ aplicables a cada funcionalidad de un PW.

### 3.4. Validación

Para validar el conjunto de atributos obtenidos en la etapa anterior se desarrolló una encuesta. El objetivo de este estudio fue conocer la importancia que un grupo de consumidores de datos le asignaba a cada atributo de DQ del modelo. Como resultado se eliminó el atributo *Información de la fuente* y sólo 33 de los 34 atributos iniciales fueron mantenidos en el modelo teórico final (ver Tabla 2). Mayores detalles acerca del desarrollo del modelo teórico pueden encontrarse en [4].

## 4. Modelo Operacional

Una vez definido el modelo teórico, la siguiente fase consistió en convertir PDQM en un modelo operacional.

Accesibilidad	Especialización	Organización
Actualidad	Exactitud	Relevancia
Aplicabilidad	Expiración	Representación
		Concisa
Atractivo	Facilidad de operación	Representación
	Flexibilidad	Consistente
Cantidad de Datos	Fiabilidad	Reputación
Complejidad	Interactividad	Seguridad
Credibilidad		Soporte de Usuario
Disponibilidad	Interpretabilidad	Tempo de Respuesta
		Trazabilidad
Documentación	Novedad	Validez
Duplicidad	Objetividad	Valor agregado
Entendibilidad	Oportunidad	

Tabla 2. Conjunto final de Atributos de DQ de PDQM

Para esto, básicamente se requiere: definir una estructura para organizar los atributos de DQ y, dada esta estructura, seleccionar/definir medidas para la evaluación de la DQ.

Para la definición de la estructura del modelo hemos tomado en cuenta una serie de propiedades o requisitos que debería satisfacer la versión final de PDQM. Entre las más importantes podemos mencionar que PDQM debe ser: *Genérico* (aplicable en cualquier tipo de PW), *Adecuado* (orientado al punto de vista del consumidor de datos, por tanto, debe soportar la subjetividad e incertidumbre asociada a la evaluación de la DQ), *Flexible* (aplicable en diferentes situaciones, por ejemplo: en diferentes dominios de PW o en procesos de evaluación donde el modelo pueda ser aplicado en forma parcial o completa) y *Completo* (la estructura debe permitir representar todas las relaciones entre los atributos).

Por otro lado, en la definición del modelo operacional se ha considerado la doble naturaleza (objetiva y subjetiva) de la calidad de datos y sus consecuencias en la forma que ésta debe ser medida. Esto es, ésta no sólo debe ser calculada con la ayuda de indicadores generados automáticamente, sino también debe ser evaluada de acuerdo con los juicios y opiniones (subjetivos) de los consumidores de datos [6].

Considerando todo lo anterior, para la operacionalización de PDQM se decidió utilizar el enfoque probabilístico, basado en redes bayesianas, propuesto en [16].

Una BN es un gráfico acíclico dirigido cuyos nodos representan variables y los arcos representan relaciones de dependencia entre las variables. Los arcos de una BN conectan a nodos padres y nodos hijos, donde la distribución de probabilidad de un nodo hijo es condicional a la distribución de probabilidad de su nodo padre. Los arcos, nodos y probabilidades pueden ser

definidos por expertos y/o derivados de datos empíricos. Las BN modelan problemas que involucran incertidumbre, y combinan las ventajas de una representación visual intuitiva con la base matemática de las probabilidades bayesianas [20]. En nuestro contexto las BN ofrecen un interesante framework con el cual es posible:

- Representar interrelaciones entre atributos de DQ de una manera intuitiva y explícita mediante la conexión de todos aquellos atributos que representan una influencia.
- Manejar los problemas de incertidumbre asociados a la subjetividad.
- Usar efectivamente la red obtenida para predecir/estimar la DQ de un portal Web.
- Aislar los factores responsables, cuando se tiene un bajo nivel de DQ.

En resumen, el proceso de transformación de PDQM se centra en la definición de una BN que lo represente y en la definición de las medidas necesarias para que en conjunto con la BN, PDQM pueda evaluar la DQ de un PW. A continuación se describirá cada una de las etapas desarrolladas para obtener el modelo operacional.

#### 4.1. Etapa 1: Un criterio para organizar los atributos de DQ

Como se explica en [20], una BN puede ser construida a partir de unidades semánticamente significativas llamadas fragmentos de red. Un fragmento es un conjunto de variables aleatorias, que pueden ser construidos y razonados por separado de otros fragmentos.

En consecuencia la definición del criterio para organizar los atributos de PDQM se ha orientado a la posibilidad de crear fragmentos. Por ello hemos decidido usar el framework de DQ propuesto en [22], que identifica cuatro categorías de DQ (Intrínseca, Accesibilidad, Contextual y de Representación) las que han sido adaptadas a nuestro contexto (redefiniendo Accesibilidad como Operacional) y en las cuales hemos procedido a clasificar los atributos de DQ (ver Tabla 3).

Esta clasificación fue hecha considerando la literatura estudiada, las definiciones de los atributos y categorías y muestras percepciones y experiencia. Como resultado, se han generado cuatro fragmentos (uno por cada categoría de DQ) que en conjunto representan una BN de tres niveles.

Categoría de DQ	Atributos de DQ
Intrínseca: Depone que los datos tienen calidad por sí mismos.	Exactitud, Objetividad, Credibilidad, Actualidad, Duplicidad, Expiración y Trazabilidad.
Operacional: Enfatiza la importancia del rol de los sistemas en la DQ. Esto es, a través de aspectos como accesibilidad, seguridad, personalización y colaboración, entre otros.	Accesibilidad, Seguridad, Interactividad, Disponibilidad, Soporte de usuario, Facilidad de operación, Tempo de respuesta.
Contextual: Esta destaca la necesidad de que la DQ sea considerada en el contexto de la tarea que se desarrolla.	Aplicabilidad, Complejidad, Flexibilidad, Novedad, Fiabilidad, Relevancia, Especialización, Oportunidad, Validez, Valor agregado.
De Representación: Denota que el sistema debe presentar los datos de manera que éstos sean interpretables y fáciles de entender, así como representados concisamente y consistentemente.	Interpretabilidad, Entendibilidad, Representación Concisa, Representación Consistente, Cantidad de Datos, Atractivo, Documentación, Organización.

Tabla 3. Cuatro categorías de DQ como criterio para organizar los atributos de PDQM.

#### 4.2. Etapa 2: Una estructura para PDQM

Una vez clasificados los atributos en cada categoría, se procedió a determinar las relaciones de dependencia entre los atributos de cada categoría. Las relaciones de dependencia o influencia directa se establecieron en forma intuitiva y en base al análisis de las definiciones de los atributos en cada categoría. A modo de ejemplo en la Tabla 4 se muestran las relaciones establecidas para la categoría de DQ de Representación.

Relaciones de Influencia Directa Nivel 2	Premisa que soporta la relación
Representación Concisa	Si los datos son representados en forma compacta, sin elementos superfluos, entonces estarán mejor representados.
Representación Consistente	Si los datos se presentan siempre con el mismo formato, y consistente con otros fuentes, entonces estarán mejor representados.
Entendibilidad	Si los datos son presentados en un lenguaje y en unidades apropiadas para la capacidad del usuario, entonces serán más entendibles.
Cantidad de datos	Si la cantidad de datos entregados por el portal es apropiada entonces serán más entendibles.
Documentación	Si los datos poseen metainformación (útil) serán más entendibles.
Organización	Si los datos están organizados con una combinación consistente de características visuales serán más entendibles.
Atractivo	Si los datos están organizados con una combinación consistente de características visuales serán más atractivos.

Tabla 4. Relaciones de DQ de Representación

Con las relaciones establecidas en todas las categorías, se generó el gráfico de la BN que representa a PDQM. Como muestra la Figura 3, en

la BN creada se pueden distinguir 4 niveles. Nivel 0, donde PDQ es el nodo que representa la DQ de todo el PW. Nivel 1, donde los nodos representan a cada categoría de DQ (obviamente el nodo PDQ es definido en término de estos cuatro nodos). Nivel 2, donde los nodos representan a los atributos que tienen influencia directa sobre cada una de las categorías de DQ. Y finalmente, el nivel 3, donde los nodos representan a los atributos que tienen influencia directa sobre los nodos del nivel 2.

En la siguiente etapa, la BN debe ser preparada para ser usada para evaluar la DQ. Esto significa que deben ser definidas las tablas de probabilidades para los nodos intermedios (nodos con padres) y las medidas para los nodos de entrada (nodos sin padres).

#### 4.3. Etapa 3: Preparación de PDQM para la evaluación de DQ

Considerando el tamaño de la BN y teniendo la posibilidad de trabajar independientemente con cada uno de los fragmentos (o subredes) generados, decidimos desarrollar esta etapa por separado para cada fragmento. Por cada fragmento se deberán desarrollar las siguientes sub-etapas:

- Si fuera necesario, crear nodos artificiales para simplificar el fragmento, esto es, reducir el número de padres por nodo.
- Definir variables cuantificables por cada nodo de entrada en el fragmento.
- Definir la tablas de probabilidad de cada nodo.

Una consideración adicional en esta etapa fue que la evaluación de la DQ es dependiente del contexto [14]. Esto significa que las demandas y necesidades de los consumidores de datos variarán dependiendo del dominio particular del PW a evaluar. Por tanto, para reflejar esto la sub-etapa variará de un dominio a otro.

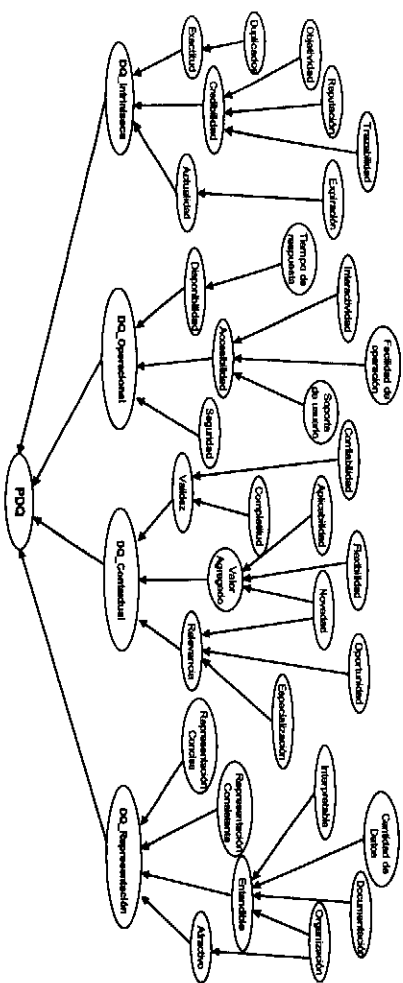


Figura 3. Red bayesiana que representa a PDQM

A continuación se presenta la preparación del fragmento DQ de Representación considerando el dominio de portales universitarios.

**a. Creación de nodos Artificiales.**

Como se observa en la Figura 4, el fragmento original tenía dos nodos con 4 padres (Entendible y DQ de Representación) y para simplificar la sub-red se crearon dos nodos artificiales: *Representación* y *Voluntad de Datos*. Con esto se evitara la explosión combinatoria al momento de generar las tablas de probabilidades de los nodos.

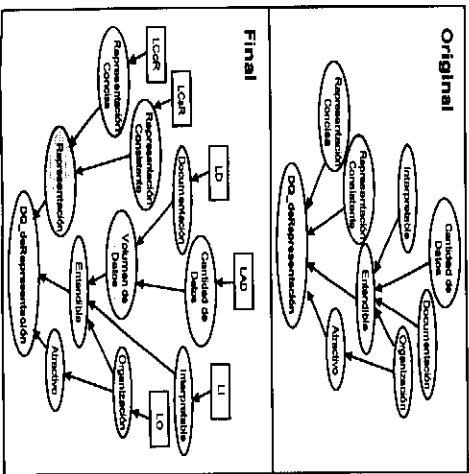


Figura 4. Creación de nodos artificiales

**b. Definición de variables cuantificables.**

En esta sub-etapa se definieron variables cuantificables (indicadores) para cada uno de los nodos de entrada en el fragmento (rectángulos en el último nivel del fragmento final en la Figura 4).

Para calcular los indicadores se seleccionaron y definieron medidas de acuerdo con la definición de cada atributo de DQ representado por los nodos de entrada. Para calcular cada indicador se siguieron 2 métodos: (1) usar medidas base y derivadas cuando se podía medir objetivamente (una medida derivada cuando se calcula a partir otras medidas base o derivadas [1]) o (2) usar valuaciones de consumidores de datos cuando el atributo no puede ser medido objetivamente. En ambos casos los indicadores tomarán un valor numérico entre 0 y 1. Como el número de posibles valores para cada nodo de entrada puede ser infinito, ha sido necesario transformarlos en variables discretas. De acuerdo con [16], esto se puede lograr usando lógica difusa. Entonces, por cada indicador definimos una función de membresía que transforma el valor obtenido en un conjunto de probabilidades asociado a las etiquetas/clases definidas.

A modo de ejemplo, a continuación explicamos la definición del indicador LCSR (Level of Consistent Representation) que genera la medida para el nodo Representación Consistente. Las medidas seleccionadas para este atributo se centran en la consistencia de formatos y en la compatibilidad de las páginas del PW. Esto, no sólo porque estos aspectos son percibidos más fácilmente por los usuarios sino porque, además, ellos pueden ser medidos en forma objetiva. Entonces, se definieron medidas basadas en el uso de estilos de páginas en el PW y también en la correspondencia entre páginas de origen y páginas de destino. Por ejemplo, un tipo de correspon-

dencia medida fue la de si el texto asociado con un enlace era repetido en la página destino, una manera simple de consistencia de datos.

Entonces para el indicador LCSR se definió un *Modelo de análisis* [1], que incluye una *fórmula* que nos da un valor numérico entre 0 y 1, y un *criterio de decisión*, en la forma de una función de membresía, que nos permite transformar el valor generado por la fórmula en un conjunto de probabilidades para las etiquetas difusas asociadas al nodo de entrada. La tabla 5 muestra un resumen de este modelo de análisis.

<b>Fórmula</b>	$LCSR = PSSD * 0.5 + SIDCD * 0.5$
<b>Medidas derivadas:</b>	PSSD: Páginas con el mismo estilo. SIDCD: Correspondencia entre fuente y destino
<b>Criterio de Decisión</b>	

Tabla 5. Modelo de análisis del indicador LCSR

**c. Definición de tablas de probabilidades**

Por cada nodo del fragmento se definieron tablas de probabilidades. Las probabilidades definidas para los nodos de entrada son derivadas directamente de los valores tomados por el correspondiente indicador. Los nodos intermedios son aquellos definidos por sus padres y que no son medidos directamente, las tablas de probabilidades de estos nodos fueron generadas por expertos, quienes las generaron tomando en cuenta el contexto de los portales universitarios.

Siguiendo con el ejemplo anterior, en la tabla 6, se muestra la tabla de probabilidades generadas para el nodo Representación Consistente.

LCSR	Low	Medium	High
Bad	0.8	0.08	0.04
Medium	0.12	0.8	0.14
Good	0.08	0.12	0.82

Tabla 6. Tabla de probabilidades del nodo Representación Consistente.

Una vez definidas todas las tablas de probabilidades, el fragmento está preparado para efectuar la evaluación de DQ de Representación en un

portal universitario dado.

Retomando el ejemplo anterior, el proceso de evaluación trabajaría de la siguiente forma cuando se calcula el indicador LCSR. Dado un portal universitario a evaluar, en cada una de sus páginas se calculan las medidas base y derivadas definidas, luego a partir de éstas se calculan las medidas derivadas PSSD y SIDCD. A partir de éstas se calcula la fórmula definida para el indicador LCSR, obteniendo un valor numérico (entre 0 y 1). Usando el criterio de decisión, el valor obtenido es transformado en un conjunto de probabilidades para las etiquetas "Low", "Medium" y "High". Con estos valores, y usando las probabilidades mostradas en la tabla 5, se derivan los valores para el nodo *Representación Consistente*. A continuación estos datos, considerados como "evidencia", serán propagados vía enlaces causales (relaciones de dependencia) hacia los nodos hijos del nodo *Representación Consistente*, en este caso hacia el nodo *Representación*, ver figura 4.

Este proceso es similar para el resto de los nodos de entrada de la BN. En general, la idea es que los valores de los nodos de entrada son medidos directamente, los valores obtenidos son transformados en un conjunto de probabilidades (correspondientes a las etiquetas asociadas a cada nodo). Estos valores de entrada son conocidos como evidencia, la cual se propagada al resto de la BN mediante enlaces causales hasta calcular el último valor que sería el correspondiente al nodo DQ de Representación.

**4.4. Validación de PDQM**

Al igual que la etapa anterior, la validación sólo fue desarrollada para el fragmento de DQ de Representación. La validación consistió en aplicar sobre un determinado PW universitario dos métodos alternativos de evaluación, uno con valuaciones directas de un grupo de consumidores de datos y otro usando PDQM. Luego comparamos los resultados obtenidos para determinar el grado de coincidencia de PDQM con la evaluación hecha por los consumidores de datos. Esto es, el grado en que PDQM representa la perspectiva de éstos.

La primera estrategia utilizada fue el desarrollo de un experimento a través del cual obtuvimos las valoraciones de un grupo de sujetos respecto de la



DQ de representación de los datos de un determinado PW universitario. Los sujetos que participaron en el experimento eran un grupo de estudiantes universitarios. El grupo estaba compuesto por 79 sujetos, todos con experiencia previa como consumidores de datos en PW. El material experimental consistió en un documento que incluía las instrucciones y motivaciones, la URL del PW, tres actividades a desarrollar en el PW y un conjunto de 9 preguntas en las cuales se pedían valoraciones de la DQ de Representación del PW. Las ocho primeras preguntas pedían valoraciones sobre cada uno de los atributos de DQ del fragmento y la última pedía una valoración global de la DQ de Representación. Los resultados de este experimento son mostrados en la tabla 8, columna sujetos.

La segunda estrategia consistió en evaluar la DQ de Representación usando PDQM. Para esto usamos la herramienta PDQA cuyo objetivo es implementar PDQM (es decir calcular las medidas definidas y aplicar la BN para obtener el nivel de la DQ), aunque por el momento es un prototipo que soporta sólo el fragmento de DQ de

Representación. A través de PDQA se puede evaluar automáticamente la DQ de Representación de un PW dado. Por razones de espacio, en este artículo sólo comentaremos los resultados obtenidos con ella y no detalles de su implementación o funcionamiento. Aplicando PDQA sobre el mismo PW evaluado con el experimento obtuvimos los siguientes valores calculados para cada indicador, (ver la Tabla 7).

LCsR	LCsR	LD	LAD	LI	LO
0.12	0.99	0.46	0.99	0.5	0.44

Tabla 7. Resultado de los indicadores

Estos valores fueron transformados en entradas válidas para la BN e ingresados como evidencia en cada nodo de entrada. Mediante la propagación de esta evidencia la BN calculó los valores finales, como se puede ver en la Figura 5.

Como una forma más concreta de comparar los valores obtenidos a través de las dos estrategias de medición usadas, en la tabla 8 se resumen los valores obtenidos.

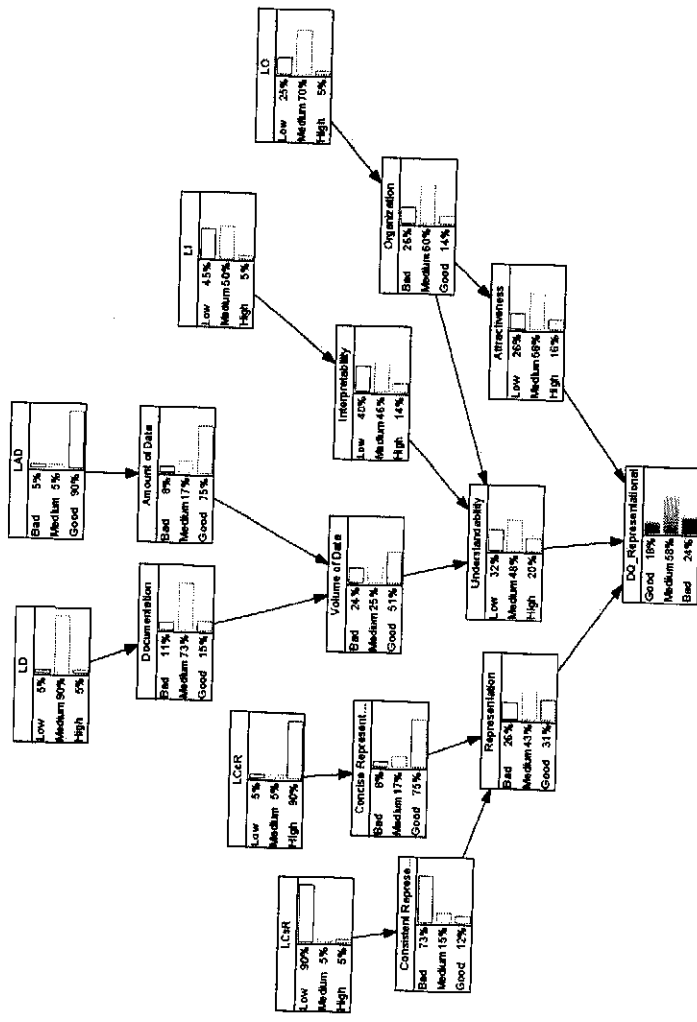


Figura 5. Resultados calculados por la BN de DQ de Representación

modelo PDQM, cuyo objetivo es la evaluación de la DQ en portales web usando la perspectiva del consumidor de datos.

Para el desarrollo de PDQM primero hemos creado un modelo teórico y luego, usando un enfoque probabilístico basado en redes bayesianas, lo hemos operativizado. Con esto pretendemos obtener un modelo final que entre otros requisitos cumpla los de ser: *Genérico* (aplicable en cualquier tipo de PW), *Adecuado* (orientado al punto de vista del consumidor de datos), *Flexible* (aplicable en diferentes situaciones) y *Completo* (que represente todas las relaciones entre los atributos).

Creemos que nuestra propuesta representa una buena alternativa para los consumidores de datos. Incluso, puede ser útil para los desarrolladores que deseen saber si su PW tiene un buen nivel de DQ para los consumidores de datos.

Como trabajo futuro realizaremos nuevas validaciones del fragmento de DQ de Representación incluyendo un mayor número de PW y de sujetos. Junto con esto, ampliaremos la definición de PDQM a otros dominios de PW. Finalmente, planeamos completar el desarrollo de los otros fragmentos de PDQM e incluirlos en PDQA.

**Agradecimientos**

Este trabajo es parte de los siguientes proyectos: ESPINGE (TIC2006-15175-C05-05) financiado por la Dirección General de Investigación del Ministerio de Ciencia y Tecnología (Spain), CALPISO (TIN20005-24055-E) financiado por el Ministerio de Educación y Ciencia (Spain), DIMENSIONES (PBC-05-012-1) financiado por FEDER y por la "Consejería de Educación y Ciencia, Junta de Comunidades de Castilla-La Mancha" (Spain) y COMPETISOFT (506AC0287) financiado por CYTED.

**Referencias**

- [1] Bertoa, M., García, F., y Vallecillo, A. An Ontology for Software Measurement, in Software Technology, Calero, C., Ruiz, F., y Piattini, M., Eds., (2006).
- [2] Bouzeghoub, M. y Peralta, V. A Framework for Analysis of data Freshness, International

Tabla 8. Resultado de ambas estrategias de evaluación

Atributo Evaluado	Valoraciones			
	Low/Bad Subj. PDQA	Medium Subj. PDQA	High/Good Subj. PDQA	High/Good Subj. PDQA
Atractivo	30%	26%	16%	28%
Organización	37%	26%	44%	60%
Cant. de Datos	18%	8%	49%	17%
Entendible	32%	32%	47%	48%
Interpretable	6%	40%	45%	46%
Documentación	16%	11%	49%	73%
R. Consistente	18%	73%	53%	15%
R. Concisa	16%	8%	52%	17%
Portal	17%	18%	68%	58%

Al observar los resultados obtenidos a nivel del portal podemos decir, en general, que los resultados entregados por PDQM son similares a los de los sujetos del experimento. Sin embargo, a pesar que ambas evaluaciones generaron como resultado que el nivel de DQ es *Medio*, es necesario observar que no existe coincidencia total en los valores obtenidos (ver última fila en Tabla 7). Además, los valores parciales tampoco coinciden e incluso algunos son muy diferentes (como R. Concisa para la valoración Low y Cantidad de datos para la valoración High). Nosotros creemos que esta diferencia se debe a los valores extremos obtenidos para los indicadores y, por tanto, será necesario revisar sus definiciones y la forma de medirlos.

Además de esto, creemos que otra razón para estas diferencias puede ser la calidad de diseño del PW evaluado. Por ejemplo, para calcular el indicador LAD (Level of Amount of Data), es necesario conocer la distribución de palabras por página. El portal medido presenta valores para esta medida que pueden ser considerados como valores atípicos. Esto significa que estos valores deberán ser eliminados del cálculo de las medidas y, por tanto, PDQA debe ser ajustada.

Junto con lo anterior, creemos que es necesario realizar nuevos experimentos en los cuales se aumente el número de sujetos y también el número de portales evaluados. La idea es poder ajustar el modelo y, por supuesto, también PDQA para que sea capaz de estimar con mayor exactitud la DQ de Representación en cualquier PW.

**5. Conclusiones y Trabajo Futuro**

En este artículo se ha presentado el desarrollo del

- Workshop on Information Quality in Information Systems, (IQIS2004), Paris, France, (2004), pp. 59-67.
- [3] Cappiello, C., Francalanci, C., y Pernici, B. Data quality assessment from the user's perspective. International Workshop on Information Quality in Information Systems, (IQIS2004), Paris, Francia, (2004), pp. 68-73.
- [4] Caro, A., Calero, C., Caballero, I., y Piattini, M. A Theoretical Model for Portal Data Quality Data Consumer Oriented, Data & Knowledge Engineering, Sent for publishing, (2007).
- [5] Collins, H. Corporate Portal Definition and Features: AMACOM, (2001).
- [6] Eppler, M. Managing Information Quality: Increasing the Value of Information in Knowledge-intensive Products and Processes: Springer, (2003).
- [7] Eppler, M., Algesheimer, R., y Dimpfel, M. Quality Criteria of Content-Driven Websites and Their Influence on Customer Satisfaction and Loyalty: An Empirical Test of an Information Quality Framework, 8th International Conference on Information Quality, (2003), pp. 108-120.
- [8] Fugini, M., Mecella, M., Plebani, P., Pernici, B., y Scannapieco, M. Data Quality in Cooperative Web Information Systems, 2002.
- [9] Gertz, M., Ozsu, T., Saake, G., y Sartler, K.-U. Report on the Dagstuhl Seminar "Data Quality on the Web", SIGMOD Record, vol. 33, Nº 1, (2004), pp. 127-132.
- [10] Graefe, G. Incredible Information on the Internet: Biased Information Provision and a Lack of Credibility as a Cause of Insufficient Information Quality, 8th International Conference on Information Quality, (2003), pp. 133-146.
- [11] Katerattanakul, P. y Siau, K. Measuring Information Quality of Web Sites: Development of an Instrument, 20th International Conference on Information System, (1999), pp. 279-285.
- [12] Katerattanakul, P. y Siau, K. Information quality in internet commerce desing, in Information and Database Quality, Piattini, M., Calero, C., y Genero, M., Eds.: Kluwer Academic Publishers, (2001).
- [13] Kitchenharn, B. Procedures for Performing Systematic Reviews, 0400011T.1, (2004).
- [14] Knight, S. A. y Burn, J. M. Developing a Framework for Assessing Information Quality on the World Wide Web, Informing Science Journal, 8, (2005), pp. 159-172.
- [15] Mahdavi, M., Shepherd, J., y Benattallah, B. A Collaborative Approach for Caching Dynamic Data in Portal Applications, 15th conference on Australian database, (2004), pp. 181-188.
- [16] Malak, G., Sahrouri, H., Badri, L., y Badri, M. Modeling Web-Based Applications Quality: A Probabilistic Approach, 7th International Conference on Web Information Systems Engineering, Wuhan, China, (2006), pp. 398-404.
- [17] Melkas, H. Analyzing Information Quality in Virtual service Networks with Qualitative Interview Data, 9th International Conference on Information Quality, (2004), pp. 74-88.
- [18] Moustakis, V., Litos, C., Dalivigas, A., y Tsironis, I. Website Quality Assessment Criteria, 9th International Conference on Information Quality, (2004), pp. 59-73.
- [19] Naumann, F. y Rolker, C. Assessment Methods for Information Quality Criteria, 5th International Conference on Information Quality, (2000), pp. 148-162.
- [20] Neil, M., Fenton, N. E., y Nielsen, L. Building large-scale Bayesian Networks, The Knowledge Engineering Review, 15(3), (2000), pp. 257-284.
- [21] Redman, T. Data Quality: The field guide. Boston: Digital Press, (2000).
- [22] Strong, D., Lee, Y., y Wang, R. Data Quality in Context, Communications of the ACM, Vol. 40, Nº 5, (1997), pp. 103-110.
- [23] Tayi, G. y Redman, T. Examining Data Quality, Communications of the ACM, Vol. 41, Nº 2, (1998), pp. 54-57.
- [24] Xiao, L. y Dasgupta, S. User Satisfaction with Web Portals: An empirical Study, in In Web Systems Design and Online Consumer Behavior, Gao, Y., Ed.: Idea Group Publishing, Hershey, PA, (2005), pp. 193-205.
- [25] Yang, Z., Cai, S., Zhou, Z., y Zhou, N. Development and validation of an instrument to measure user perceived service quality of information presenting Web portals, Information and Management Elsevier Science, 42, (2004), pp. 575-589.