

**Conferência iadis Ibero-Americana**



# **WWW/Internet 2007**

**7 - 8 Outubro  
Vila Real, Portugal**

## **Actas**

**Editado por:  
Ramiro Gonçalves  
Flavia Maria Santoro  
Pedro Isaías  
José María Gutiérrez**



**iadis**

international association for development of the information society

**CONFERÊNCIA IADIS IBERO-AMERICANA**

**WWW/INTERNET 2007**



**ACTAS DA CONFERÊNCIA IADIS IBERO-AMERICANA**  
**WWW/INTERNET 2007**

**VILA REAL, PORTUGAL**

**OUTUBRO, 7 - 8, 2007**

Organizado por  
**IADIS**

**International Association for Development of the Information Society**

Co-Organizado por  
**Universidade de Trás-os-Montes e Alto Douro**

Copyright 2007

IADIS Press

Todos os direitos reservados

Este trabalho está sujeito a direitos de autor. Todos os direitos são reservados, no todo ou em parte, mais especificamente os direitos de tradução, reimpressão, reutilização de ilustrações, re-citação, emissão, reprodução em microfilme ou de qualquer outra forma, e armazenamento em bases de dados. A permissão para utilização deverá ser sempre obtida da IADIS Press. Por favor contactar [secretariat@iadis.org](mailto:secretariat@iadis.org).

Editado por Ramiro Gonçalves, Flavia Maria Santoro, Pedro Isaías e José María Gutiérrez

Editores Associados: Luís Rodrigues e Patrícia Barbosa

ISBN: 978-972-8924-45-4

# ÍNDICE

PREFÁCIO	xi
COMITÉ DO PROGRAMA	xiii
PALESTRA	xix

## ARTIGOS LONGOS

EM BUSCA DE UM ENSINO COMPLEMENTAR E DE QUALIDADE: O PROJETO DESAFIO DIGITAL <i>Eduardo Morgado, Maria Ângela Dias dos Santos, Barbara De Franco e Daniel Igarashi da Cruz</i>	3
GERAÇÃO (SEMI)AUTOMÁTICA DE METADADOS: UM CONTRIBUTO PARA A RECUPERAÇÃO DE OBJECTOS DE APRENDIZAGEM <i>Vitor Gonçalves e Eurico Carrapatoso</i>	10
MAPEAMENTO E ANÁLISE DAS TROCAS INTERINDIVIDUAIS EM AMBIENTES DE EDUCAÇÃO A DISTÂNCIA <i>Patrícia B. Scherer Bassani e Patricia Behar</i>	20
UMA FERRAMENTA PARA ACOMPANHAMENTO DE ACESSOS A CONTEÚDOS DIDÁTICOS <i>Christiane Meiler Baptista, Regina Melo Silveira e Wilson Vicente Ruggiero</i>	27
RECOMENDACIÓN DE PERFILES ACADÉMICOS MEDIANTE ALGORITMOS COLABORATIVOS BASADOS EN EL EXPEDIENTE <i>Emilio J. Castellano, Luís Martínez, Manuel Barranco y Luis G. Pérez Cordón</i>	35
UMA FERRAMENTA DE MINERAÇÃO DE TEXTOS PARA ASSISTIR AS DÚVIDAS DOS ALUNOS EM AMBIENTES VIRTUAIS DE APRENDIZAGEM <i>Eros Estevão de Moura, Sahudy Montenegro González, Annabell del Real Tamariz e Geórgia R. Rodrigues Gomes</i>	43
HIPERMÍDIA ADAPTATIVA COMO ESTRATÉGIA APLICADA AO E-COMMERCE <i>Oscar Dalfovo, Pedro S. Zanchett e Jefferson W. Mette</i>	51
MODELO PARA LA GESTIÓN DEL E-LEARNING CORPORATIVO EN EL AMBITO UNIVERSITARIO <i>Eugenio Fernández</i>	59
A ATENÇÃO AO ECONÓMICO, AO SOCIAL E AO AMBIENTE ATRAVÉS DAS PÁGINAS INTERNET DO PODER LOCAL <i>Luís Manuel Ribeiro Vieira</i>	67

E-GOV BRASILEIRO: UMA ANÁLISE DOS PORTAIS DOS MUNICÍPIOS BRASILEIROS <i>Ademir Martinez Sanches, Marcelo Augusto Santos Turine and Débora Maria Barroso Paiva</i>	75
MSB: UNA APLICACIÓN WWW PARA GENERACIÓN DE RESÚMENES DE COMPORTAMIENTO <i>Víctor Flores y Martín Molina</i>	83
DISEÑO DE UNA HERRAMIENTA PARA EL DESARROLLO DE APLICACIONES WEB BASADAS EN STRUTS <i>Adelaide Bianchini, Ricardo Blanch, Maruja Ortega y Ascánder Suárez</i>	91
UM MODELO ONTOLÓGICO PARA O CONTEXTO TURÍSTICO <i>Salvador Lima e José Moreira</i>	99
INTEGRACIÓN SEMÁNTICA DE INFORMACIÓN PARA LA IDENTIFICACIÓN DE PERSONAS: UN ENFOQUE CON TECNOLOGÍAS DE LA WEB SEMÁNTICA <i>René F. Navarro-Hernández, César E. Rose Gómez, Gilberto Gradías Enríquez y José A. Pacheco Sánchez</i>	107
EN BUSCA DEL TIPO DE SERVICIO WEB SEMÁNTICO MÁS APROPIADO PARA DAR SOPORTE A LA GRID SEMÁNTICA <i>Jorge Martinez-Gil, Adolfo Lozano-Tello y Antonio Polo-Marquez</i>	115
DETECCIÓN Y SEGUIMIENTO DE INTRUSOS EN REDES BLUETOOTH CORPORATIVAS <i>Ramiro Cano y Luis Bengochea</i>	122
USABILIDAD EN UN JUEGO DE MEMORAMA CON RECONOCIMIENTO DE VOZ PARA NIÑOS <i>Carlos Miranda-Palma, Reyna Camal-Uc, José Cen-Magaña, Cinhtia Gonzalez-Segura, Sergio Gonzalez-Segura, Michel García-García y Lizzie Narvaez-Díaz</i>	129
DEFINIENDO UNA ESTRUCTURA DE EVALUACIÓN PARA MEDIR LA USABILIDAD DE SITIOS WEB EDUCATIVOS <i>María E. Alva Obeso, Ana B. Martínez Prieto, María del C. Suárez Torrente, Hernán Sagástegui Chigne, y Juan M. Cueva Lovelle</i>	137
EXPERIÊNCIA NA CUSTOMIZAÇÃO DE UM MODELO DE PROCESSO DE SOFTWARE PARA DESENVOLVIMENTO DE APLICATIVO PARA WEB UTILIZANDO OO <i>Marla Teresinha Barbosa Geller e Rodrigo Quites Reis</i>	145
ALGORITMOS PARA TOKENS DE AUTENTICAÇÃO <i>Gustavo Yamasaki Martins Vieira e Wilson Vicente Ruggiero</i>	153
ÉTICA, PRIVACIDADE E SEGURANÇA EM SISTEMAS HIPERMÍDIA ADAPTATIVOS PARA EDUCAÇÃO BASEADA EM WEB <i>Sandra Gavioli Puga e Maria Alice Grigas Varella Ferreira</i>	160
CRITÉRIOS PARA A AVALIAÇÃO DA QUALIDADE DOS RECURSOS E DA INFORMAÇÃO DISPONÍVEL NA INTERNET <i>Catarina Costa e Bráulio Alturas</i>	168
EVOLUCIÓN DE LA TOPOLOGÍA DE INTERNET A NIVEL DE SISTEMAS AUTÓNOMOS BASADO EN EL PROYECTO ROUTEVIEWS <i>José Luis Gahete Díaz, Natalia Fernández Gallego, Ana María Garzo Ortega y Gonzalo Martín Villaescusa</i>	175

INTERNET COMO CANAL DE COMPRA: RESPUESTAS DE CONSUMO ANTE DIFERENTES ESTÍMULOS WEB	183
<i>Carlota Lorenzo Romero, Alejandro Mollá Descals y Miguel Ángel Gómez Borja</i>	
USO DA INTERNET NO TURISMO: UMA PESQUISA EMPÍRICA COM TURISTAS INTERNACIONAIS EM UM DESTINO NO BRASIL	191
<i>Leilianne Michelle Trindade da Silva, Mabel Simone de Araújo Bezerra Guardia, Anatólia Saraiva Martins Ramos e Sérgio Ramiro Rivero Guardia</i>	
SIMULADOR DO SERVIDOR DE HABILITAÇÃO DE LOCALIZAÇÃO EM SISTEMA LBS	199
<i>Leonardo Batista de Queiroz e Fernanda Maria Ribeiro Alencar</i>	
ANALISE DE DESEMPENHO DE PROTOCOLOS DE CRIPTOGRAFIA EM REDES SEM FIO	207
<i>Camila Suzin, Walter Priesnitz Filho e Maria Emilia Camargo</i>	
AAUTOMATOR: HERRAMIENTA FLEXIBLE PARA LA EXTRACCIÓN DE INFORMACIÓN EN SITIOS WEB BIOINFORMÁTICOS	215
<i>Daniel Glez-Peña, José R. Méndez, Florentino Fdez-Riverola</i>	
MODELO DE QUALIDADE DE COMPONENTES DE SOFTWARE	223
<i>Regina Maria Thienne Colombo, Ana Cervigni Guerra, Maria Teresa Villalobos Aguayo e Darley Rosa Peres</i>	
UTILIZACIÓN DE GOOGLE SEARCH APPLIANCE PARA LA BÚSQUEDA Y PLANIFICACIÓN DE INFORMACIÓN GEOGRÁFICA A TRAVÉS DE UN PROTOTIPO Y SU POSIBLE UTILIZACIÓN EN DISPOSITIVOS GPS	231
<i>Rubén Gonzalez-Crespo, Javier Parra-Fuente, Oscar Sanjuan-Martínez y Luis Joyanes-Aguilar</i>	

## **ARTIGOS CURTOS**

EMOÇÕES COMO PARTE DE UM AMBIENTE VIRTUAL DE APRENDIZAGEM	243
<i>Annabell D.R. Tamariz, Sahudy Montenegro González, Cinthia B. Pessanha e Edson C. Braga</i>	
ADAPTACIÓN DE UN LMS OPEN SOURCE. APLICACIÓN DE CONTROL ESTADÍSTICO PARA MOODLE	248
<i>Miguel Ángel Conde, Carlos Muñoz, Alberto Velasco y Francisco José García</i>	
CONTEXTUALIZAÇÃO DA APRENDIZAGEM DA PROGRAMAÇÃO: ESTUDO EXPLORATÓRIO NO SECOND LIFE	253
<i>Micaela Esteves; Ricardo Antunes, Leonel Morgado; Paulo Martins e Benjamim Fonseca</i>	
ESPECIFICAÇÃO DE PERSONAS PARA EDUCAÇÃO BASEADA NA WEB COM FOCO NOS ESTILOS DE APRENDIZAGEM DO ALUNO	258
<i>Sandra Gavioli Puga e Maria Alice Grigas Varella Ferreira</i>	
AVANCE: UM AMBIENTE DE ENSINO E APRENDIZAGEM BASEADO NA WEB 2.0	262
<i>Maurício Nunes da Costa Bomfim, João Sérgio dos Santos Assis e Fabio Ferrentini Sampaio</i>	



APLICACIÓN DE UN ÍNDICE DE EVALUACIÓN DE CALIDAD EN SITIOS WEB DE GOBIERNOS LOCALES EN ESPAÑA <i>Antonio Paños Álvarez</i>	266
GESTÃO DO ACOMPANHAMENTO DE ACTIVIDADES DE ALUNOS EM MUNDOS VIRTUAIS: ESTUDO EXPLORATÓRIO NO SECOND LIFE® <i>Ricardo Antunes, Leonel Morgado, Paulo Martins e Benjamim Fonseca</i>	272
PROJECTO DE APLICAÇÃO EM INTERNET PARA APOIO À DECISÃO NA REGA POR GRAVIDADE <i>André Pereira Muga e José Manuel Gonçalves</i>	277
AQUISIÇÃO E TRATAMENTO DE INFORMAÇÃO DE LOCALIZAÇÃO PARA SUPORTE A APLICAÇÕES SENSÍVEIS AO CONTEXTO <i>Edgar Silva, Marco Liberato, Hugo Paredes, Alberto Vasconcelos, António Costa, João Barroso e Paulo Martins</i>	281
UMA ARQUITECTURA PARA A COMPOSIÇÃO DINÂMICA DE SERVIÇOS DEPENDENTES DO CONTEXTO <i>João Paulo Sousa, Eurico Carrapatoso e Benjamim Fonseca</i>	286
DISEÑO, IMPLEMENTACIÓN Y PRUEBA DE UN PROTOCOLO DE TRANSFERENCIA DE DATOS <i>Marco-Antonio Balderas Cepeda, Manuel Aguirre Bortoni y Martín Vogel Vázquez</i>	290
APLICANDO CONCEPTOS DE TEORÍA DE LA INFORMACIÓN PARA EL FILTRADO DE CORREO SPAM <i>José R. Méndez, Ignacio Cid, Daniel Glez-Peña y Florentino Fdez-Riverola</i>	295
EXTRACCIÓN DE PERFILES BASADA EN AGRUPAMIENTO GENETICO PARA RECOMENDACIÓN DE CONTENIDO <i>Sandra Patricia Tocarruncho Tocarruncho, Fredy Andrés Aponte Novoa y Arturo Tocarruncho Tocarruncho</i>	299
A UTILIZAÇÃO DE TECNOLOGIAS NO PROCESSO CRIATIVO ARQUITETÔNICO: UMA ANÁLISE SOBRE AS SUAS INFLUÊNCIAS E POSSÍVEIS MUDANÇAS <i>Luciana Dalfollo Ferreira, Paulo Horn Regal e Leda Lísia Franciosi Portal</i>	304
PORTAIS WEB: ENQUADRAMENTO CONCEPTUAL <i>António Jorge Gonçalves de Gouveia, Paula Cristina Oliveira e João Eduardo Quintela Varajão</i>	309
SOCIALIZACIÓN DE LA WEB SEMÁNTICA <i>Jorge Morato, Anabel Fraga, Sonia Sánchez-Cuadrado y Jose Antonio Moreira</i>	315
UMA IMPLEMENTAÇÃO DE BAM - BUSINESS ACTIVITY MONITORING NO GOVERNO BRASILEIRO <i>Fernando José Travassos Vieira</i>	320
NOVOS DESAFIOS E OPORTUNIDADES DE INVESTIGAÇÃO NA ÁREA DA GESTÃO DE PROJECTOS DE DESENVOLVIMENTO DE SISTEMAS DE INFORMAÇÃO <i>Dulce Gonçalves, Rui Rijo, Ramiro Gonçalves, José Bulas Cruz e João Varajão</i>	324
MODELOS DE MATURIDADE DO COMÉRCIO ELECTRÓNICO <i>Elisabete Paulo Morais, Ramiro Gonçalves e José Adriano Pires</i>	330
PANORÁMICA Y SITUACIÓN DEL ESTÁNDAR EBXML <i>Mario Antonio Triguero y Roberto Barchino</i>	335

REENGENHARIA DE UM SISTEMA DE INFORMAÇÃO DE APOIO AO ENSINO UNIVERSITÁRIO PARA A CRIAÇÃO DE RESULTADOS ESTATÍSTICOS <i>Pedro Branco, Sónia Santos, Luís Barbosa e Ramiro Gonçalves</i>	340
MATERIALIZAÇÃO DE VANTAGENS COMPETITIVAS MEDIANTE A ADOÇÃO DE SISTEMAS ERP <i>Ana Carolina Sousa de Almeida, Ana Flávia Silva Simões, Denise de Andrade Ribeiro e Sílvio Vanderlei Araújo Sousa</i>	345
CATEGORIZAÇÃO AUTOMÁTICA DE DOCUMENTOS PDF PARA BIBLIOTECAS DIGITAIS <i>Geórgia R. Rodrigues Gomes, Igor Primo Curttis, Erlon Márcio Couto Alves, Sahudy Montenegro González e Annabell Del Real Tamariz</i>	351
INTERFACES PARA APRENDIZAGEM DE TAREFAS COLABORATIVAS ESPACIAIS <i>Filipe Santos, Leonel Morgado, Paulo Martins e Benjamim Fonseca</i>	356
UMA ARQUITETURA TOLERANTE A FALHAS PARA PROCESSOS DE NEGÓCIO BASEADOS EM SERVIÇOS WEB <i>Diego Zuquim Guimarães Garcia e Maria Beatriz Felgar de Toledo</i>	361
CLASIFICACIÓN DE DOCUMENTOS XML SEGÚN CRITERIOS DE CALIDAD DE DATOS <i>Eugenio Verbo, Ismael Caballero, Coral Calero y Mario Piattini</i>	366
UMA PROPOSTA PARA O DESENVOLVIMENTO DE MATERIAL DIDÁTICO BASEADO NO PERFIL DO USUÁRIO <i>Patrícia B. Scherer Bassani, Marine Bello Flores, Marcelo Ritzel e Regina de Oliveira Heidrich</i>	371

## **ARTIGOS DE REFLEXÃO**

MERCHANDISING EN INTERNET: NUEVAS TENDENCIAS DEL MARKETING EN EL PUNTO DE VENTA <i>Carlota Lorenzo Romero, Juan Antonio Mondéjar Jiménez y José Mondéjar Jiménez</i>	379
E_REINAC: ELEARNING REUSABILIDAD, INTEROPERATIVIDAD Y ACCESIBILIDAD WEB: ASPECTOS TECNOPEDAGÓGICOS <i>Mariluz Cacheiro González, Covadonga Rodrigo San Juan y Miguel Rodríguez Artacho</i>	383
BIBLIOTECAS DIGITALES: CONCEPTOS <i>Gomes, N. F</i>	387
ESTRATEGIA COLABORATIVA Y DE CONOCIMIENTO PARA LA IGUALDAD DE GÉNERO <i>Fernando Giner</i>	391
E-MERCADOS MEDIANTE P2P: UN NUEVO MARCO PARA LAS PYMES <i>Josef Ignacio Hötz Ordoño y M<sup>a</sup> Teresa Ariza Gómez</i>	395
O CONTRIBUTO DAS TIC PARA O DESENVOLVIMENTO SUSTENTADO DO PATRIMÓNIO <i>Natália Botica, Luís Fontes e Ana Roriz</i>	399

## POSTERS

UNA ONTOLOGÍA DE VINOS ESPAÑOLES <i>Óscar Gil, Ana- Belén Gil y Francisco José García</i>	407
FACTORES CRÍTICOS DA ADESÃO DAS PME'S NACIONAIS, FORNECEDORAS DE MATERIAIS DE ESCRITÓRIO AO PROCEDIMENTO AQUISITIVO PÚBLICO EM PORTUGAL: UMA PROPOSTA DE INVESTIGAÇÃO <i>Paulo Pereira e Bráulio Alturas</i>	411
REPOSITÓRIO CIENTÍFICO DIGITAL – UTAD <i>Orlando Carvalho, Guilherme Saraiva, Jorge Godinho, Hugo Paredes e João Barroso</i>	414
ESPECIFICAÇÃO DE UMA APLICAÇÃO PARA A CRIAÇÃO E MANUTENÇÃO DE CURRÍCULOS VITAE USANDO TÉCNICAS UML <i>Ana Rodrigues, Marlene Pinto, João Varajão e Ramiro Gonçalves</i>	417
AGENTES INTELIGENTES NO AMBIENTE VIRTUAL DE ENSINO DE LÓGICA HALYEN <i>Sahudy Montenegro González, Annabell del Real Tamariz, Eduardo Coelho Carneiro e Jamile Souza de Almeida</i>	421
ANÁLISIS DE LOS OBJETIVOS DE NEGOCIO, ESTRATEGIAS DE ÉXITO Y BARRERAS EN SITIOS WEB DE EMPRESAS ESPAÑOLAS <i>Antonio Paños Álvarez</i>	424
ANÁLISE DE PROJECTOS DE INVESTIMENTO NA ÁREA DAS TECNOLOGIAS DA INFORMAÇÃO: MÉTODOS E TÉCNICAS <i>Ricardo Zenida e Bráulio Alturas</i>	427
PORTAL ERASMUS <i>João Gonçalves, Victor Matos, Hugo Paredes, António Marques, Paulo Martins e Benjamim Fonseca</i>	430
UTILIZAÇÃO DE OPÇÕES REAIS NA ANÁLISE DE PROJECTOS DE INVESTIMENTO EM S.I./T.I.C. <i>António Braz e Bráulio Alturas</i>	433
SISTEMA DE GESTÃO DE PROJECTOS ACADÉMICOS <i>João Braz Pinto, Ana Pinto de Matos, Hugo Paredes, António Marques, Paulo Martins e João Paulo Moura</i>	437
USO DA TÉCNICA DE CSP NO DESENVOLVIMENTO DE SISTEMAS DE INFORMAÇÃO PARA LOGÍSTICA DE DISTRIBUIÇÃO DE MERCADORIAS <i>Oscar Dalfovo, Paulo Maurício Selig e Rodrigo Kammer</i>	440
USO DE REDES NEURAIIS NO DESENVOLVIMENTO DE SISTEMAS APLICAÇÃO NAS EMPRESAS DO SETOR TÊXTIL NA ÁREA FINANCEIRA PARA APROVAÇÃO DO LIMITE DE CRÉDITO <i>Oscar Dalfovo, Paulo Maurício Selig e Edson Vander Souza</i>	444

## AUTHOR INDEX

# PREFÁCIO

Estas actas contêm os artigos e posters da Conferência IADIS Ibero-Americana WWW/Internet 2007, organizada pela *International Association for Development of the Information Society* e co-organizada pela Universidade de Trás-os-Montes e Alto Douro, em Vila Real, Portugal, Outubro, 7 e 8, 2007.

O objectivo desta conferência é de constituir um fórum para reunir investigadores, alunos e demais pessoas que trabalhem ou investiguem o campo da WWW/Internet. As seguintes cinquenta e duas áreas foram objecto de submissões de artigos e posters:

Acessibilidade; Sistemas Web Adaptáveis; Colaboração; Comunicação Mediada por Computador; Data Mining; Planeamento e Desenvolvimento de Bases de Dados; Bibliotecas Digitais e Publicação Electrónica; Aplicações Distribuídas e Paralelas; E-Business e E-Commerce; E-Government; E-Learning; EDI; Avaliação da Qualidade; Linguagens Extensíveis; Tendências Globais da WWW/Internet; Trabalho em Grupo; Interação Homem-Máquina; Hipermedia; Arquitecturas de Informação; Visualização de Informação; Agentes Inteligentes; Interfaces; Serviços Internet; Linguagens; Metadados; Multimédia; Aspectos de Desempenho; Personalização de Serviços e Sítios da Web; Estratégias em Portais; Protocolos e Standards; Pesquisa e Navegação; Aspectos de Segurança; Semântica da Web; Armazenamento; Integração de Sistemas; Estratégias de Ensino e Aprendizagem; Inovação e Competitividade Tecnológica; Administração Tecnológica; Estratégias Tecnológicas; Tele-Trabalho; Aplicações WWW/Internet; Casos de Estudo; Impacto da WWW/Internet; Engenharia Web; Personalização da Web; Aplicações Wireless; Computação Ubíqua; Usabilidade; Modelação do Utilizador; Comunidades Virtuais; Realidade Virtual; XML.

A conferência Ibero-Americana IADIS WWW/Internet 2007 teve cerca de 145 submissões. Cada submissão foi avaliada por uma média de três revisores independentes para assegurar o elevado nível final das submissões aceites. O resultado final foi a publicação de 30 artigos longos, sendo publicados também artigos curtos, artigos de reflexão e posters.

Como sabemos, a organização de uma conferência requer o esforço de muitas pessoas. Gostaríamos de agradecer a todos os membros do Comité de Programa pelo trabalho realizado na revisão e selecção dos artigos que constam destas actas.

Estas actas resultam da contribuição de um variado número de autores. Estamos gratos a todos os autores que submeteram os seus artigos. Também gostaríamos de agradecer ao Professor Doutor Carlos Ramos, Director do GECAD, ISEP, Instituto Politécnico do Porto, Portugal, por ter aceite dar uma palestra. Também gostaríamos de agradecer a todos os membros do comité de organização, delegados, e convidados cuja contribuição e envolvimento são cruciais para o sucesso desta conferência.

Por fim desejamos que todos os participantes tenham uma excelente estadia em Vila Real.

José María Gutiérrez, Escuela Politécnica de la Universidad de Alcalá, Espanha

Ramiro Gonçalves, Universidade de Trás-os-Montes e Alto Douro, Portugal

Flavia Maria Santoro, Universidade Federal do Estado do Rio de Janeiro, Brasil

Pedro Isaías, Universidade Aberta, Portugal

*Chairs*

Vila Real, Portugal

Outubro 2007

# COMITÉ DO PROGRAMA

## CO-CHAIRS

Ramiro Gonçalves, Universidade de Trás-os-Montes e Alto Douro, Portugal  
Flavia Maria Santoro, Universidade Federal do Estado do Rio de Janeiro, Brasil  
Pedro Isaías, Universidade Aberta, Portugal

## PROGRAM CHAIR

José María Gutiérrez, Escuela Politécnica de la Universidad de Alcalá, Espanha

## MEMBROS DO COMITÉ

Adelaide Bianchini, Universidad Simón Bolívar, Venezuela  
Adolfo Duran, Universidade Federal da Bahia, Brasil  
Adolfo Lozano, University of Extremadura, Espanha  
Afonso Inacio Orth, PUCRS, Brasil  
Alejandro Gonzalez, National University of La Plata, Argentina  
Alejandro Zunino, Tandil University, Argentina  
Alberto Raposo, PUC-Rio, Brasil  
Alfonso Rodríguez, Universidad De Castilla-la Mancha, Espanha  
Álvaro Suárez Sarmiento, Las Palmas de Gran Canaria University, Espanha  
Ana Ayerbe Fernandez-Cuesta, Robotiker Tecnalia, Espanha  
Angélica de Antonio, Universidad Politécnica de Madrid, Espanha  
Annabell del Real Tamariz, Universidade Estadual do Norte Fluminense, Brasil  
Antonio Gabriel López Herrera, Universidad de Jaén, Espanha  
Antonio Paños Alvarez, Universidad de Murcia, Espanha  
Aqueo Kamada, Centro de Pesquisas Renato Archer, Brasil  
Arnaldo Belchior, Universidade de Fortaleza, Brasil  
Benjamim Fonseca, Universidade de Trás-os-Montes e Alto Douro, Portugal  
Carlos Andrés Romano, Universidad Politécnica de Valencia, Espanha  
Carlos Costa, ISCTE, Portugal

Carlos Juiz, Universidad de las Islas Baleares, Espanha  
Carlos Porcel Gallego, Universidad de Córdoba, Espanha  
Christina Chavez, DCC/UFBA, Brasil  
Claudia Lage Rebello da Motta, Universidade Federal Do Rio De Janeiro, Brasil  
Clevi Elena Rapkiewicz, Universidade do Norte Fluminense, Brasil  
Clifton E. Clunie B., Universidad Tecnológica de Panamá, Panamá  
Coral Calero, Universidad De Castilla-la Mancha, Espanha  
Cristiana Bentes, Universidade do Estado do Rio de Janeiro, Brasil  
Cristina Barrado, Universitat Politècnica de Catalunya, Espanha  
Cristina Cachero, Universidad de Alicante, Espanha  
Cristóbal Romero Morales, Universidad de Córdoba, Espanha  
Daniel Fernández Lanvin, Universidad de Oviedo, Espanha  
Daniel Gayo Avello, Universidad de Oviedo, Espanha  
Débora Conforto, UFRGS, Brasil  
Denis Silveira, Faculdades Ibmecc/RJ, Brasil  
Eliane Loiola, COPPE/UFRJ, Brasil  
Elsa M<sup>a</sup> Macías-López, Las Palmas de Gran Canaria University, Espanha  
Emilio Insfran, Universidad Politécnica de Valencia, Espanha  
Enrique Herrera Viedma, Universidad de Granada, Espanha  
Eva Gibaja Galindo, Universidad de Córdoba, Espanha  
Eva Lorenzo Iglesias, Universidad de Vigo, Espanha  
Fábio Ferrentini Sampaio, NCE – UFRJ, Brasil  
Fabrício Silva, Universidade Católica de Santos, Brasil  
Federico Botella, Universidad Miguel Hernández, Espanha  
Fernanda Alencar, Universidade Federal de Pernambuco, Brasil  
Fernanda Baião, UNIRIO, Brasil  
Fernanda Campos, UFJF, Brasil  
Florentino Fernandez Riverola, Universidad de Vigo, España  
Gabriela Arevalo, National University Of La Plata, Argentina  
Georgia Gomes, Universidade Cândido Mendes-Campos, Brasil  
Gonzalo Méndez, Universidad Complutense de Madrid, Espanha

Gustavo Rossi, National University Of La Plata, Argentina  
Hermes Senger, Universidade Católica de Santos, Brasil  
Isabel Harb Manssour, PUCRS, Brasil  
Isabel Sofia Brito, Escola Superior de Tecnologia e Gestão de Beja, Portugal  
Jaime Ramírez, Universidad Politécnica de Madrid, Espanha  
Jesús Rodríguez Perez, Capgemini, Espanha  
João Garrott, ISEGI-UNL / Universidade Lusofona, Portugal  
João Varajão, Universidade de Trás-os-Montes e Alto Douro, Portugal  
Jorge Barbosa, UNISINOS, Brasil  
Jorge Sá Silva, Universidade de Coimbra, Portugal  
José Antonio Delgado Osuna, Universidad de Córdoba, Espanha  
José Eduardo Córcoles Tintero, Universidad de Castilla-La Mancha, Espanha  
José Maria Nazar David, Faculdade Ruy Barbosa, Brasil  
José Ramón Méndez Reboredo, University of Vigo, Espanha  
José Raúl Romero, Universidad de Córdoba, Espanha  
Juan Carlos Fernández Caballero, Universidad de Córdoba, Espanha  
Juan López Rubio, Escola Politecnica Superior de Castelldefels, Espanha  
Juan Manuel Pérez, Universitat Jaume I, Espanha  
Katja Gilly, Universidad Miguel Hernández, Espanha  
Liliana Passerino, UFRGS, Brasil  
Liliane dos Santos Machado, Universidade Federal da Paraíba, Brasil  
Lucia Giraffa, PUCRS, Brasil  
Luís Guerrero, Universidad De Chile, Chile  
Marcelo Turine, Universidade Federal do Mato Grosso do Sul, Brasil  
Marco Aurélio Gerosa, PUC-Rio, Brasil  
Marco Painho, Universidade Nova de Lisboa, Portugal  
Marcos Borges, UNICAM, Brasil  
Marco Winckler, IRIT, France  
Maria Angélica Caro Gutierrez, Universidad De Castilla-la Mancha, Espanha  
Maria Clícia Stelling De Castro, Universidade Do Estado Do Rio De Janeiro, Brasil  
M<sup>a</sup> del Puerto Paule Ruiz, Universidad de Oviedo, Espanha



María Dolores Ayuso, Universidad de Murcia, Espanha  
Maria Lencastre, Universidade de Pernambuco, Brasil  
María Luque Rodríguez, Universidad de Córdoba, Espanha  
Mariano Pimentel, PUC-Rio, Brasil  
Mário Freire, Universidade da Beira Interior, Portugal  
Marta González, Robotiker Tecnalia, Espanha  
Mauro Pequeno, Universidade Federal Do Ceará, Brasil  
M. Cecilia C. Baranauskas, Universidade Estadual de Campinas, Brasil  
Miguel de Castro Neto, Universidade Nova de Lisboa, Portugal  
Nathalie Moreno, Universidad de Málaga, Espanha  
Olga Revilla, Itákora, Espanha  
Olga Santos Martín, UNED, Espanha  
Oscar Von Pamel, Universidad Nacional de Rosario, Argentina  
Patricia Alejandra Behar, Universidade Federal Do Rio Grande Do Sul, Brasil  
Patrícia Scherer Bassani, Centro Universitário Feevale, Brasil  
Paula Oliveira, Universidade de Trás-os-Montes e Alto Douro, Portugal  
Paulo Martins, Universidade de Trás-os-Montes e Alto Douro, Portugal  
Paulo Simões, Universidade de Coimbra, Portugal  
Raúl Antonio Aguilar Vera, Universidad Autónoma de Yucatán, Mexico  
Regina Heidrich, Centro Universitário Feevale, Brasil  
Regina Lúcia de Oliveira Moraes, UNICAM, Brasil  
Renata Mendes de Araújo, Universidade do Rio de Janeiro, Brasil  
Ricardo Farias, Universidade Federal do Rio de Janeiro, Brasil  
Ricardo Llamasa-Villalba, Universidad Industrial de Santander, Colombia  
Rita Suzana Pitangueira Maciel, Faculdade Ruy Barbosa, Brasil  
Rodrigo Bonacin, Centro de Pesquisas Renato Archer, Brasil  
Rosa M. Vicari, CINTED/UFRGS, Brasil  
Rosa Rita Maenza, Universidad Tecnológica Nacional, Argentina  
Sahudy Montenegro Gonzalez, Universidade Estadual do Norte Fluminense, Brasil  
Salvador Alcaraz, Universidad Miguel Hernández, Espanha  
Sean Siqueira, UNIRIO, Brasil

Sebastián Lozano, Universidad de Sevilla, Espanha  
Sebastian Ventura Soto, Universidad de Córdoba, Espanha  
Sérgio Crespo, Universidade do Vale do Rio dos Sinos, Brasil  
Silvia Baldiris, University of Girona, Espanha  
Sílvia Abrahão, Universidad Politécnica De Valencia, Espanha  
Simone Bacelar Leal Ferreira, UNIRIO, Brasil  
Simone Diniz Junqueira Barbosa, PUC-Rio, Brasil  
Susana Marchisio, Universidad Nacional de Rosario, Argentina  
Susana Torrado Morales, Universidad de Murcia, Espanha  
Vitor Basto Fernandes, Universidade de Trás-os-Montes e Alto Douro, Portugal  
Vicente Luque Centeno, Universidad Carlos III de Madrid, Espanha  
Victor Manuel Ruiz Penichet, Universidad de Castilla-La Mancha, Espanha

# CLASIFICACIÓN DE DOCUMENTOS XML SEGÚN CRITERIOS DE CALIDAD DE DATOS

Eugenio Verbo, Ismael Caballero  
*Departamento de Investigación y Desarrollo*  
*Indra Software Factory, SLU*  
*Centro Mixto de Investigación y Desarrollo UCLM-Indra*  
*Ronda de Toledo s/n – 13004 Ciudad Real, España*  
*{emverbo, icaballerom}@indra.es*

Coral Calero, Mario Piattini  
*Grupo de Investigación Alarcos*  
*Departamento de Tecnologías y Sistemas de Información*  
*Centro Mixto de Investigación y Desarrollo UCLM-Indra*  
*Paseo de la Universidad 4 – 13071 Ciudad Real, España*  
*{Coral.Calero, Mario.Piattini}@uclm.es*

## RESUMEN

En los últimos años, XML se ha distinguido como la tecnología preferida para el intercambio de datos. Potenciada por las Arquitecturas Orientadas a Servicios (SOA), su uso ha permitido simplificar el acceso y distribución de fuentes heterogéneas de datos. Si bien una de sus grandes ventajas es el alto grado de interoperabilidad que permite, esta característica se puede convertir en un inconveniente en caso de que se acceda de forma incontrolada a las fuentes de datos. Para disminuir este riesgo es necesario aplicar principios de calidad de datos para valorar los orígenes de datos y seleccionar los que mejor se adapten a los requisitos de los usuarios que utilizan el sistema. Nuestra propuesta consiste en la aplicación de técnicas de softcomputing para la valoración de características de las fuentes de datos y, en base a ellas, clasificarlas de acuerdo a una jerarquía de criterios predeterminada. De esta forma, sólo los documentos XML que superen el umbral de calidad de datos requerido serán usados por el sistema, mejorando la calidad del producto final.

## PALABRAS CLAVES

XML, datos semiestructurados, calidad de datos, lógica difusa.

## 1. INTRODUCCIÓN

La popularización de las Arquitecturas Orientadas a Servicios ha propiciado el auge de XML como tecnología de intercambio de datos para conseguir la máxima interoperabilidad entre sistemas distribuidos y heterogéneos. El éxito de este tipo de sistemas depende fuertemente de la calidad de los datos que manejen. En (Strong et al., 1997) se exponen y justifican los diez principales tipos de problemas relacionados con la falta de calidad de datos. Dos de ellos se dan principalmente en sistemas distribuidos: a) múltiples orígenes del mismo dato producen diferentes valores y b) sistemas distribuidos heterogéneos conducen a la inconsistencia de definiciones, formatos y valores. Ello motiva la necesidad de contar con documentos XML que tengan la suficiente calidad en los datos que se almacenan y comparten. Muchas veces los documentos vienen de fuentes con las que ya se ha tratado, por lo que puede ser posible predecir la calidad que van a tener los datos que lleguen. Para ello se hace imprescindible disponer de técnicas y herramientas que permitan una clasificación automática de un documento XML antes de ser utilizado. Esta clasificación se basará en una serie de reglas que deberán ser inferidas tras la evaluación de numerosos documentos provenientes de la misma fuente. Nuestra contribución consiste en una metodología que permita la elaboración de semejante sistema de reglas para la clasificación de documentos XML según los criterios o dimensiones de calidad de datos elegidos por las diferentes partes interesadas. Para establecer este sistema de reglas se necesitan medidas cuantitativas y cualitativas sobre la calidad de datos que tiene un documento

XML. Para la elaboración de este sistema de reglas se puede necesitar de la opinión de muchos expertos. Dado que estas opiniones son subjetivas, se introduce la imprecisión. Para gestionar estas imprecisiones se debe utilizar la lógica difusa, que permite la agregación de diferentes opiniones.

En muchas ocasiones, para poder tener valoraciones de datos almacenados en un documento XML es necesario aportar una serie de metadatos que complementen el significado de los datos con respecto a la dimensión de calidad que cada parte interesada desempeñando un rol específico necesite evaluar. En (Verbo et al., 2007) se propone una estructura de datos basada en XML Schema, DQXSD, para almacenar conjuntamente los datos junto con sus correspondientes metadatos.

El resto del artículo sigue la siguiente estructura: en la sección 2 se hace una pequeña introducción al concepto de calidad de datos, en la sección 3 se explica el enfoque propuesto para la clasificación difusa de documentos XML, y en la sección 4 se presentan una serie de conclusiones y líneas de trabajo futuro.

## 2. CONCEPTO Y DIMENSIONES DE CALIDAD DE DATOS

La definición más aceptada de calidad de datos es “*adecuación para su uso*” (“*fitness for use*”): los datos tienen calidad si sirven para el propósito que se necesita. A la vista de esta definición, se deduce que la calidad de datos es un concepto subjetivo (depende del uso que vaya a hacer cada persona) y multidimensional (es factible entender la calidad desde distintos puntos de vista – al estilo del software con ISO 9126). A los diferentes puntos de vista se les llama dimensiones de calidad de datos. Estas dimensiones son ampliamente discutidas y explicadas en la literatura (Lee et al., 2006).

Para poder hacer una gestión adecuada de la calidad de los datos, es necesario definir medidas para cada una de las dimensiones requeridas para un escenario. La forma en la que se definen las medidas depende fuertemente del método de medición que cada parte interesada necesite emplear según sus necesidades: así una misma dimensión de calidad puede tener tantas medidas diferentes como partes interesadas haya en un escenario. En cualquier caso, las medidas de calidad de los conjuntos de datos pueden clasificarse como orientadas al modelo de datos (suelen ser objetivas) o las orientadas a los propios datos (pueden requerir metadatos que complementen el significado del dato con respecto a la dimensión de calidad de datos según la parte interesada). En (Lee et al., 2006) y en (Batini and Scannapieco, 2006) se dan indicaciones oportunas sobre cómo definir estas medidas. Estas medidas serán la base para la creación del sistema de reglas.

## 3. CLASIFICACIÓN DIFUSA DE DOCUMENTOS XML

El objetivo de nuestra propuesta es la construcción de un sistema que permita clasificar documentos XML de acuerdo a la calidad de los datos que contiene. Puesto que la evaluación de la calidad de datos es un proceso muy dependiente del dominio de aplicación y de una subjetividad inherente, debe ser capaz de manejar imprecisión y ser configurable para cada contexto en el que se quiera implantar. Las técnicas de softcomputing dan soporte a estas necesidades ya que la imprecisión es uno de sus principios básicos.

En concreto, para nuestra propuesta nos hemos basado en la técnica de tablas de repositorios difusos, propuesta en (Castro-Sánchez et al., 2001) y que ha sido aplicada con éxito en distintos dominios como la adquisición de conocimiento y sistemas de ayuda a la decisión. Esta técnica se basa en la utilización de opiniones de expertos para extraer las características definitorias de los documentos que están siendo clasificados. Después de varias iteraciones y mediante el uso de la tabla de repositorios difusos, explicada más adelante, se obtiene un sistema de reglas que permite diferenciar los documentos utilizados como entrada del sistema. En nuestro enfoque, estos documentos serán ficheros XML sobre los que se realizarán una serie de mediciones tras las que, después de aplicar el sistema de reglas obtenido anteriormente, se podrá devolver como resultado el nivel de calidad de datos asignado al documento XML.

El proceso de clasificación se divide en las etapas que se describen a continuación. Para explicar mejor este proceso, en cada uno de estos pasos se dará un ejemplo teórico:

1. Identificar los niveles de calidad en los que se desean dividir los documentos XML, usando por ejemplo una escala Likert en la que se podrían dividir en documentos de calidad “*alta*”, “*media*” y “*baja*”.

2. Proporcionar casos de estudio controlados para el aprendizaje del sistema de reglas. Ello requiere formar un conjunto de documentos XML de muestra en el que haya al menos dos documentos para cada nivel de calidad. El nivel de calidad de datos de estos documentos se establecerá de acuerdo a la experiencia previa de los expertos implicados en el proceso, de forma que el uso de los documentos haya permitido revelar su utilidad para el fin requerido. En nuestro ejemplo teórico, se podría tener un conjunto de muestra formado por seis elementos: dos documentos de alta calidad,  $D_1$  y  $D_2$ ; dos de calidad media,  $D_3$  y  $D_4$ ; y dos de baja calidad,  $D_5$  y  $D_6$ .
3. Seleccionar un subconjunto aleatorio de tres elementos de los elementos del conjunto de muestra que todavía no hayan sido distinguidos. Por ejemplo,  $\{D_1, D_4, D_5\}$ .
4. Identificar una dimensión de calidad de datos común a los tres documentos del subconjunto que permita distinguir cada uno de ellos, es decir, que el valor de la medida de la dimensión sea distinto para cada documento. En (Castro-Sánchez et al., 2001) se propone que esta decisión quede a juicio de un experto en el dominio de aplicación. No obstante no deja de ser un juicio subjetivo difícil de revisar. Para disminuir el grado de subjetividad, se propone el uso de algunas de las medidas referidas en (Verbo et al., 2007) o. Sin embargo, no se puede obviar la alta dependencia respecto al dominio de aplicación que tiene el criterio a utilizar. Así, por ejemplo, una noticia deportiva puede ser de gran relevancia para un sistema que procese información deportiva, pero tener importancia nula para un sistema de análisis bursátil. De esta forma, también resulta necesario introducir en este paso características dependientes del contexto como, por ejemplo, el grado de fiabilidad de una determinada fuente de datos.

Como resultado se obtiene la tabla de valores para esa característica, que muestra sus distintos valores difusos. En la Tabla 1 se pueden ver los rangos de valores definidos para la medida GCD (bajo, medio y alto) mientras que en la Figura 1 se puede ver su representación gráfica. La medida GCD se utiliza para calcular el grado de completitud de un documento y su definición, que queda fuera del alcance de este artículo, puede ser consultada en (Verbo et al., 2007) junto con las definiciones de otras medidas interesantes para el propósito de esta técnica.

Tabla 1. Valores de la variable GCD

	Bajo	Medio	Alto
a	0	0,5	0,8
b	0	0,6	0,9
c	0,5	0,8	-
d	0,6	0,9	-

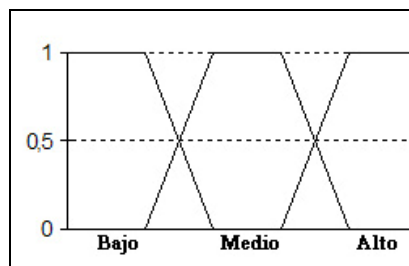


Figura 1. Representación gráfica de los valores de la variable GCD

Los valores de  $a$ ,  $b$ ,  $c$  y  $d$  definen la función trapezoidal habitualmente utilizada para representar valores difusos. Siendo  $v$  el valor continuo de la variable a calificar, esta función trapezoidal se define de la siguiente forma:

$$\Pi(v; a, b, c, d) = \begin{cases} 0 & v < a, \\ \frac{(v-a)}{(b-a)} & a \leq v \leq b, \\ 1 & b \leq v \leq c, \\ \frac{(d-v)}{(d-c)} & c \leq v \leq d, \\ 0 & v > d. \end{cases} \quad (1)$$

5. Aplicar las medidas sobre cada uno de los documentos del conjunto de muestra y, en función del resultado obtenido, asignarles un valor difuso para la característica escogida. Este resultado se almacena en la tabla del repositorio difuso (Tabla 2).

Tabla 2. Tabla del repositorio difuso tras la introducción de la primera variable

	<b>D<sub>1</sub></b>	<b>D<sub>2</sub></b>	<b>D<sub>3</sub></b>	<b>D<sub>4</sub></b>	<b>D<sub>5</sub></b>	<b>D<sub>6</sub></b>
GDC	Alto	Alto	Medio	Alto	Medio	Bajo

6. Calcular la distancia entre los distintos elementos del conjunto de muestra. Esta distancia consiste en una función que puede ser definida por el propio usuario o también se puede utilizar alguna otra como la propuesta en (Castro-Sánchez et al., 2001). Aquellos documentos que superen un umbral predeterminado, serán considerados diferentes. Los que no lo superen, se considerarán similares. El resultado obtenido es la matriz de distinciones, que muestra las distancias entre los resultados obtenidos para cada documento. En la Tabla 3 se puede ver un ejemplo en el que los pares de documentos {D<sub>1</sub>, D<sub>6</sub>}, {D<sub>2</sub>, D<sub>6</sub>} y {D<sub>4</sub>, D<sub>6</sub>} tienen una distancia igual a 1 que es representada en la tabla con la nomenclatura  $C_1(I)$ , es decir, los documentos de cada par son totalmente diferentes respecto a la medida utilizada que, en este caso, ha sido GCD.
7. Repetir los pasos 3 y 4 hasta que se hayan distinguido todos los documentos.
8. Extraer una serie de reglas que modelen el comportamiento del sistema. Son reglas del tipo *si-entonces*. Por ejemplo, supóngase que se tienen dos medidas, Grado de Compleción del Documento (GCD) y Profundidad Media del Árbol de Datos de Calidad (DQDT), ambas definidas en (Verbo et al., 2007), entonces se podría enunciar la siguiente regla:
  - **SI** GCD es alto y DQDT es 4, **ENTONCES** el documento es de calidad “alta”.

Tabla 3. Matriz de distinciones después de la introducción de la primera variable

	<b>D<sub>1</sub></b>	<b>D<sub>2</sub></b>	<b>D<sub>3</sub></b>	<b>D<sub>4</sub></b>	<b>D<sub>5</sub></b>	<b>D<sub>6</sub></b>
<b>D<sub>1</sub></b>	Nil	Nil	Nil	Nil	Nil	$C_1(1)$
<b>D<sub>2</sub></b>	Nil	Nil	Nil	Nil	Nil	$C_1(1)$
<b>D<sub>3</sub></b>	Nil	Nil	Nil	Nil	Nil	Nil
<b>D<sub>4</sub></b>	Nil	Nil	Nil	Nil	Nil	$C_1(1)$
<b>D<sub>5</sub></b>	Nil	Nil	Nil	Nil	Nil	Nil
<b>D<sub>6</sub></b>	$C_1(1)$	$C_1(1)$	Nil	$C_1(1)$	Nil	Nil

Derivando un conjunto de reglas se puede construir un sistema que se encargue de clasificar documentos XML, según las dimensiones de calidad de datos elegidas como las más significativas para los usuarios.

#### 4. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se ha propuesto una metodología para la clasificación automática de documentos XML según su calidad de datos asociada. Para ello, se han aplicado principios de *softcomputing* para manejar la imprecisión de las características de calidad de cada documento. Habitualmente, estas técnicas se aplican a partir del juicio de un experto. A pesar de que los conocimientos del experto en el campo de aplicación son amplios, éstos no dejan de ser subjetivos. En la solución propuesta se trata de minimizar esta subjetividad mediante el uso de medidas que se aplican sobre los documentos utilizados para diseñar el sistema. No obstante, la intervención de los expertos sigue siendo imprescindible pues es necesaria para la elaboración de los rangos difusos de las medidas.

Las líneas de trabajo futuro se dirigen en primer lugar, a la validación del algoritmo propuesto sobre documentos reales en distintos contextos de uso. Más adelante, por un lado se procederá a la identificación de las dimensiones de calidad de datos con mayor impacto en la clasificación, y, por otro, a un refinamiento de la metodología mediante su aplicación al mayor número de casos de estudio posibles. Aparte del refinamiento, se pretende determinar los límites de los conjuntos difusos usados en el proceso de clasificación. Asimismo, posterior a la obtención del conjunto de reglas iniciales, sería recomendable generalizar el conjunto de reglas resultante mediante la aplicación del algoritmo propuesto en (Castro et al., 1999).

## AGRADECIMIENTOS

Esta investigación es parte de los proyectos FAMOSO (FIT-340000-2007-71) subvencionado por el Ministerio de Industria, Turismo y Comercio, ESFINGE (TIN2006-15175-C05-05) y CALIPSO (TIN 2005-24055-E), subvencionados por el Ministerio of Educación y Ciencia.

## REFERENCIAS

- Batini, C. & Scannapieco, M. n., 2006. *Data Quality: Concepts, Methodologies and Techniques*, Berlin, Springer-Verlag Berlin Heidelberg.
- Castro-Sánchez, J. J., Castro, J. L. & Zurita, J. M., 2001. Fuzzy Repertory Table: A Method for Acquiring Knowledge About Input Variables to Machine Learning Algorithm. *IEEE Transactions on Fuzzy Systems*, 12, 123-139.
- Castro, J. L., Castro-Sánchez, J. J. & Zurita, J. M., 1999. Learning maximal structure rules in fuzzy logic for knowledge acquisition in expert systems. *Fuzzy Sets and Systems*, 101, 331-342.
- Lee, Y. W., Pipino, L. L., Funk, J. D. & Wang, R. Y., 2006. *Journey to Data Quality*, Cambridge, MA, USA, Massachussets Institute of Technology.
- Strong, D. M., Lee, Y. W. & Wang, R. Y., 1997. 10 Potholes in the Road to Information Quality. *IEEE Computer*.
- Verbo, E., Caballero, I., Fernández-Medina, E. & Piattini, M., 2007. *Data Quality in XML Databases. A Methodology for Semi-Structured Database Design Supporting Data Quality Issues*. Artículo aceptado para 2th International Conference on Software and Data Technologies (ICSOFT). Barcelona, Spain.