WEBIST 2008

FOURTH INTERNATIONAL CONFERENCE ON WEB INFORMATION SYSTEMS AND TECHNOLOGIES

Proceedings

Internet Technology
e-Learning

FUNCHAL, MADEIRA - PORTUGAL · MAY 4 - 7, 2008

CO-ORGANIZED BY

INSTICC

UNIVERSIDADE da MADEIRA

CO-SPONSORED BY

MEMBER
Workflow Management Coalition
WfMC

INSTICC PRESS

WEBIST 2008

Proceedings of the
Fourth International Conference on
Web Information Systems and Technologies

Volume 1

Funchal, Madeira, Portugal

May 4 – 7, 2008

Co-organized by
INSTICC – Institute for Systems and Technologies of Information, Control and Communication
and
UMA – Universidade da Madeira

Co-sponsored
WfMC – Workflow Management Coalition

Copyright © 2008 INSTICC – Institute for Systems and Technologies of Information, Control and Communication
All rights reserved

Edited by José Cordeiro, Joaquim Filipe and Slimane Hammoudi

Printed in Portugal

ISBN: 978-989-8111-26-5

Depósito Legal: 271876/08

http://www.webist.org

secretariat@webist.org

BRIEF CONTENTS

INVITED SPEAKERS IV

ORGANIZING AND STEERING COMMITTEES V

PROGRAM COMMITTEE VI

AUXILIARY REVIEWERS IX

SELECTED PAPERS BOOK X

FOREWORD XI

CONTENTS XIII

INVITED SPEAKERS

Tony Shan

Bank of America

U.S.A.

Leszek Maciaszek

Macquarie University

Australia

Claudia Medeiros

UNICAMP

Brazil

Marcin Paprzycki

Systems Research Institute Polish Academy of Science

Poland

Rainer Unland

University of Duisburg-Essen

Germany

Klaus Pohl

University of Duisburg-Essen

Germany

ORGANIZING AND STEERING COMMITTEES

CONFERENCE CHAIR

Joaquim Filipe, INSTICC / Polytechnic Institute of Setúbal, Portugal

PROGRAM CO-CHAIRS

José Cordeiro, INSTICC / Polytechnic Institute of Setúbal, Portugal

Slimane Hammoudi, E.S.E.O., France

LOCAL ARRANGEMENTS

Laura Rodriguez, University of Madeira (UMa), Portugal

PROCEEDINGS PRODUCTION

Andreia Costa, INSTICC, Portugal

Bruno Encarnação, INSTICC, Portugal

Helder Coelhas, INSTICC, Portugal

Paulo Brito, INSTICC, Portugal

Vera Coelho, INSTICC, Portugal

Vera Rosário, INSTICC, Portugal

Vitor Pedrosa, INSTICC, Portugal

CD-ROM PRODUCTION

Elton Mendes, INSTICC, Portugal

Pedro Varela, INSTICC, Portugal

WEBDESIGNER AND GRAPHICS PRODUCTION

Marina Carvalho, INSTICC, Portugal

SECRETARIAT AND WEBMASTER

Bruno Encarnação, INSTICC, Portugal

PROGRAM COMMITTEE

A. Y. Al-Zoubi, Princess Sumaya University for Technology, Jordan

Ajith Abraham, Norwegian University of Science and Technology, Norway

Jacky Akoka, CNAM & INT, France

Abdullah Alghamdi, KSU, Saudi Arabia

Rachid Anane, Coventry University, U.K.

Margherita Antona, FORTH - ICS, Greece

Grigoris Antoniou, FORTH-ICS, Greece

Elarbi Badidi, U.A.E University, U.A.E

Matteo Baldoni, Università degli Studi di Torino, Italy

Denilson Barbosa, University of Calgary, Canada

Cristina Baroglio, Università di Torino, Italy

Bernhard Bauer, University of Augsburg, Germany

David Bell, Brunel University, U.K.

Orlando Belo, University of Minho, Portugal

Bharat Bhargava, Purdue University, U.S.A.

Stefan Böttcher, University of Paderborn, Germany

Paolo Bouquet, University of Trento, Italy

Christos Bouras, University of Patras and RACTI, Greece

Stephane Bressan, National University of Singapore, Singapore

Amy Bruckman, Georgia Tech, U.S.A.

Maria Claudia Buzzi, CNR - IIT, Italy

Elena Calude, Massey University, IIMS, New Zealand

Nunzio Casalino, LUISS Guido Carli University, Italy

Maiga Chang, Athabasca University, Canada

Evangelos Christou, University of the Aegean, Greece

Christophe Claramunt, Naval Academy Research Institute, France

Isabelle Comyn-Wattiau, CNAM & ESSEC, France

Michel Crampes, Ecole des Mines d'Ales, France

Alexandra Cristea, University of Warwick, U.K.

Daniel Cunliffe, University of Glamorgan, U.K.

Alfredo Cuzzocrea, University of Calabria, Italy

Vasilios Darlagiannis, EPFL, Switzerland

Alessandro D'Atri, CeRSI - Luiss Guido Carli University, Italy

Valeria de Antonellis, University of Brescia, Italy

Sergio de Cesare, Brunel University, U.K.

Steven Demurjian, The University of Connecticut, U.S.A.

Darina Dicheva, Winston-Salem State University, U.S.A.

Y. Ding, University of Innsbruck, Austria

Juergen Dix, Clausthal University of Technology, Germany

Josep Domingo-Ferrer, Rovira i Virgili University of Tarragona, Spain

Chyi-Ren Dow, Feng Chia University, Taiwan

Schahram Dustdar, Vienna University of Technology, Austria

Barry Eaglestone, The University of Sheffield, U.K.

Atilla Elci, Eastern Mediterranean University, North Cyprus, Turkey

Opher Etzion, IBM, Israel

Luis Ferreira Pires, University of Twente, The Netherlands

Josep-Lluis Ferrer-Gomila, Universitat de les Illes Balears, Spain

Filomena Ferrucci, University of Salerno, Italy

Stefan Fischer, University of Luebeck, Germany

Akira Fukuda, Kyushu University, Japan

Giovanni Fulantelli, Italian National Research Council - Institute for Educational Technology, Italy

Martin Gaedke, University of Karlsruhe, Germany

Erich Gams, Salzburg Research, Austria

Daniel Garcia, Universidad de Oviedo, Spain

Franca Garzotto, HOC-Politecnico di Milano, Italy

Dragan Gasevic, Athabasca University, Canada

José A. Gil Salinas, Universidad Politécnica de Valencia, Spain

Karl Goeschka, Vienna University of Technology, Austria

PROGRAM COMMITTEE (CONT.)

Anna Grabowska, Gdansk University of Technology, Poland

Begoña Gros, University of Barcelona, Spain

Vic Grout, Centre for Applied Internet Research (CAIR), University of Wales, NEWI, U.K.

Francesco Guerra, Università di Modena e Reggio Emilia, Italy

Aaron Gulliver, University of Victoria, Canada

Pål Halvorsen, University of Oslo, Norway

Ioannis Hatzilygeroudis, University of Patras, Greece

Stylianos Hatzipanagos, King's College, London, U.K.

Dominikus Heckmann, DFKI GmbH, Germany

Nicola Henze, Leibniz University Hannover, Germany

Dominic Heutelbeck, FernUniversität in Hagen, Germany

Pascal Hitzler, AIFB, University of Karlsruhe, Germany

Wen Shyong Hsieh, Stu Te University, National Sun Yat-sen University, Taiwan

Brian Hudson, Umeå University, Sweden

Tanko Ishaya, University of Hull, U.K.

Kai Jakobs, RWTH Aachen University, Germany

Anne James, Coventry University, U.K.

Ivan Jelinek, Czech Technical University in Prague, Czech Republic

Jingwen Jin, Intel Corporation/Communications Technology Lab, U.S.A.

Qun Jin, Waseda University, Japan

Carlos Juiz, University of Balearic Islands, Spain

Michail Kalogiannakis, University Paris 5 - Rene Descartes - Laboratory: Education et Apprentissages (E.D.A.), France

Jussi Kangasharju, University of Helsinki, Finland

Vassilis Kapsalis, Industrial Systems Institute, Greece

George Karabatis, University of Maryland Baltimore County (UMBC), U.S.A.

Frank Kargl, Ulm University, Germany

Roland Kaschek, Massey University, New Zealand

Sokratis Katsikas, University of Piraeus, Greece

Hamamache Kheddouci, Université Claude Bernard Lyon 1, France

Ralf Klamma, RWTH Aachen University, Germany

Tsvi Kuflik, The University of Haifa, Israel

Axel Küpper, Ludwig Maximilian University Munich, Germany

Daniel Lemire, UQAM, Canada

Tayeb Lemloma, IUT of Lannion, France

Kin Fun Li, University of Victoria, Canada

Weigang Li, University of Brasilia, Brazil

Leszek Lilien, Western Michigan University, U.S.A.

Claudia Linnhoff-Popien, Mobile and Distributed Systems Group, Institute for Informatics, Ludwig-Maximilians-Universität Munich, Germany

Hui Liu, Missouri State University, U.S.A.

Ying Liu, Hong Kong Polytechnic University, China

Pascal Lorenz, University of Haute Alsace, France

Heiko Ludwig, IBM TJ Watson Research Center, U.S.A.

Vicente Luque-Centeno, Carlos III University of Madrid, Spain

Anna Maddalena, DISI - University of Genoa, Italy

George Magoulas, Birkbeck College, University of London, U.K.

Johannes Mayer, University of Augsburg, Germany

Ingo Melzer, Daimler AG, Germany

Elisabeth Metais, CNAM, CEDRIC Laboratory, France

Panagiotis Metaxas, Wellesley College, U.S.A.

Alessandro Micarelli, Roma Tre University, Italy

Debajyoti Mukhopadhyay, Techno India (West Bengal University of Technology), India

Kia Ng, ICSRiM - University of Leeds, U.K.

David Nichols, University of Waikato, New Zealand

Andreas Ninck, Berne University of Applied Sciences, Switzerland

Dusica Novakovic, London Metropolitan University, U.K.

Vladimir Oleshchuk, University of Agder, Norway

PROGRAM COMMITTEE (CONT.)

Andrea Omicini, Alma Mater Studiorum - Università di Bologna, Italy

Kok-Leong Ong, Deakin University, Australia

Jose A. Onieva, Universidad de Málaga, Spain

Vasile Palade, Oxford University, U.K.

Jun Pang, Oldenburg University, Germany

Kalpdrum Passi, Laurentian University, Canada

Viviana Patti, Università degli Studi di Torino, Italy

Guenther Pernul, University of Regensburg, Germany

Rajesh Pillania, Institute for Emerging Markets, India

Carsten Pils, Waterford Institute of Technology (WIT), Ireland

Pierluigi Plebani, Politecnico di Milano, Italy

Bhanu Prasad, Florida A & M University, U.S.A.

Joshua C. C, Pun, City University of Hong Kong, Hong Kong

Mimi Recker, Utah State University, U.S.A.

Frank Reichert, University of Agder, Norway

Werner Retschitzegger, Johannes Kepler University Linz, Austria

Yacine Rezgui, University of Salford, U.K.

Thomas Risse, L3S Research Center, Germany

Marco Roccetti, University of Bologna, Italy

Dumitru Roman, University of Innsbruck / DERI Innsbruck, Austria

Danguole Rutkauskiene, Kaunas University of Technology, Lithuania

Sourav S. Bhowmick, Nanyang Technological University, Singapore

Maytham Safar, Kuwait University, Kuwait

Eduardo Sanchez, Universidad de Santiago de Compostela, Spain

Abdolhossein Sarrafzadeh, Massey University, New Zealand

Anthony Savidis, ICS-FORTH, Greece

Vittorio Scarano, Università di Salerno, Italy

Alexander Schatten, Vienna University of Technology: Institute for Software Technology and Interactive Systems, Austria

Alexander Schill, TU Dresden, Germany

Jochen Seitz, Technische Universität Ilmenau, Germany

Tony Shan, Bank of America, U.S.A.

Jamshid Shanbehzadeh, Tarbiat Moallem University, Iran

Quan Z. Sheng, The University of Adelaide, Australia

Keng Siau, University of Nebraska-Lincoln, U.S.A.

Miguel-Angel Sicilia, University of Alcalá, Spain

Marianna Sigala, University of the Aegean, Greece

Eva Söderström, University of Skövde, Sweden

Pedro Soto-Acosta, University of Murcia, Spain

J. Michael Spector, Florida State University, Learning Systems Institute, U.S.A.

Martin Sperka, Slovak University of Technology, FIIT, Slovakia

Kathleen Stewart Hornsby, University of Iowa, U.S.A.

Frank Stowell, University of Portsmouth, U.K.

Carlo Strapparava, FBK-IRST, Italy

Eleni Stroulia, University of Alberta, Canada

Hussein Suleman, University of Cape Town, South Africa

Aixin Sun, Nanyang Technological University, Singapore

Junichi Suzuki, University of Massachusetts, Boston, U.S.A.

Ramayah T., Universiti Sains Malaysia, Malaysia

Taro Tezuka, Kyoto University, Japan

Dirk Thissen, RWTH Aachen University, Germany

Ivan Tomek, Acadia University, Canada

Bettina Törpel, Technical University of Denmark, Denmark

Arun Kumar Tripathi, Dresden University of Technology, Germany

Thrasyvoulos Tsiatsos, Aristotle University of Thessaloniki and Research Academic Computer Technology Institute, Greece

Klaus Turowski, University of Augsburg, Germany

Michail Vaitis, University of the Aegean, Greece

PROGRAM COMMITTEE (CONT.)

Christelle Vangenot, EPFL, Switzerland

Athanasios Vasilakos, University of Western Macedonia, Greece

Jari Veijalainen, University of Jyvaskyla, Finland

Juan D. Velasquez, University of Chile, Chile

Maria Esther Vidal, Universidad Simón Bolívar, Venezuela

Maurizio Vincini, Università di Modena e Reggio Emilia, Italy

Viacheslav Wolfengagen, Institute JurInfoR-MSU, Russian Federation

Martin Wolpers, Katholieke Universiteit Leuven, Belgium

Huahui Wu, Google, U.S.A.

Lu Yan, University College London, U.K.

Sung-Ming Yen, National Central University, Taiwan

Janette Young, Northumbria University, U.K.

Weihua Zhuang, University of Waterloo, Canada

AUXILIARY REVIEWERS

Solomon Behre, University of Connecticut, U.S.A.

Stephan Bloehdorn, University of Karlsruhe, Germany

Sebastian Blohm, University of Karlsruhe, Germany

Duygy Çelik, Eastern Mediterranean University, Turkey

Frank Fitzek, University of Agder, Norway

Andrea Forte, Georgia Tech, Georgia, U.S.A.

Vittorio Fuccella, University of Salern, Italy

Carmine Gravino, University of Salerno, Italy

Carlos Guerrero, University of the Balearic Islands, Spain

Larisa Ismailova, MEPhI, Russia

Jan Kolter, University of Regensburg, Germany

Sergey Kosikov, MEPhI, Russia

Riccardo Martoglia, Università di Modena e Reggio Emilia, Italy

Jaime A. Pavlich-Mariscal, University of Connecticut, U.S.A.

Mir Mohsen Pedram, Tarbiat Moalem University Teheran, Iran

Octavian Popescu, FBK-irst, Italy

Andreas Prinz, University of Agder, Norway

Claudio Schifanella, University of Torino, Italy

Rolf Schillinger, University of Regensburg, Germany

Francesc Sebé, Rovira i Virgili University, Spain

Daniel Serrano, University of Malaga, Spain

Mikael Snaprud, University of Agder, Norway

Agusti Solanas, Rovira i Virgili University, Spain

George Spathoulsa, University of Piraeus, Greece

Sarita Yardi, Georgia Tech, U.S.A.

Jose Zagal, Georgia Tech, U.S.A.

SELECTED PAPERS BOOK

A number of selected papers presented at WEBIST 2008 will be published by Springer-Verlag in a LNBIP Series book. This selection will be done by the Conference Chair and Program Co-chairs, among the papers actually presented at the conference, based on a rigorous review by the WEBIST 2008 program committee members.

FOREWORD

This volume contains the proceedings of the Fourth International Conference on Web Information Systems and Technologies (WEBIST 2008), co-organized by the Institute for Systems and Technologies of Information, Control and Communication (*INSTICC*) and the University of Madeira (*UMa*) and co-sponsored by the Workflow Management Coalition (*WfMC*).

The purpose of this Conference is to bring together researchers, engineers and practitioners interested in the technological advances and business applications of web-based information systems. It has four main topic areas, covering different aspects of Web Information Systems, including "Internet Technology", "Web Interfaces and Applications", "Society, e-Business, e-Government" and "e-Learning".

WEBIST 2008 received 238 paper submissions from more than 40 countries in all continents. A double-blind review process was enforced, with the help of more than 200 experts from the international program committee, all of them with a Ph.D. in one of the main conference topic areas. After reviewing, 32 papers were selected to be published and presented as full papers, i.e. completed work (8 pages in proceedings / 30' oral presentations) and 64 additional papers, describing work-in-progress as short papers for 20' oral presentation. Furthermore there were also 58 papers presented as posters. The full-paper acceptance ratio was 13%, and the total oral paper acceptance ratio was 40%. These ratios denote a high level of quality, which we intend to maintain or reinforce in the next edition of this conference.

Besides the proceedings edited by INSTICC, a post-conference book will be compiled with extended versions of its best papers, and published by Springer-Verlag. Appropriate indexing has been arranged for the proceedings of WEBIST 2008.

One of the remarkable aspects of the WEBIST conference series, since its first edition in 2005, is the expertise brought about by a large number of distinguished keynote speakers internationally recognized for their scientific level of excellence, whose lectures and whose participation in the conference panel definitely contribute to highly enhance the quality of this event. This year WEBIST hosted six keynotes, delivered by Dr. Tony Shan (Bank of America, U.S.A.), Dr. Leszek Maciaszek (Macquarie University, Australia), Dr. Claudia Medeiros (UNICAMP, Brazil), Dr. Marcin Paprzycki (Systems Research Institute Polish Academy of Science, Poland), Dr. Rainer Unland and Dr. Klaus Pohl (both from the University of Duisburg-Essen, Germany).

Building an interesting and successful program for the conference required the dedicated effort of many people. Firstly, we must thank the authors, whose research and development efforts are recorded here. Secondly, we thank the members of the program committee and additional reviewers for their diligence and expert reviewing. We also wish to include here a word of appreciation for the excellent organization provided by the conference secretariat, from INSTICC, who have smoothly prepared the most appropriate environment for a productive meeting and scientific networking. Last but not least, we thank the invited speakers for their invaluable contribution and for taking the time to synthesize and prepare their talks.

We wish you all an exciting conference and a pleasant stay in Funchal, Madeira. We hope to meet you again next year for the 5th WEBIST, in Lisbon, Portugal, details of which will be shortly made available at *http://www.webist.org*.

Joaquim Filipe
INSTICC/Polytechnic Institute of Setúbal
Portugal

José Cordeiro
INSTICC/Polytechnic Institute of Sétubal
Portugal

CONTENTS

INVITED SPEAKERS

KEYNOTE LECTURES

SOA IN PRACTICE IS-5
Tony C. Shan

SERVING ONTOLOGIES ACROSS THE WEB - Challenges and Approaches IS-7
Claudia Bauzer Medeiros

BUILDING QUALITY INTO WEB INFORMATION SYSTEMS IS-9
Leszek A. Maciaszek

GENERIC FRAMEWORK FOR AGENT ADAPTABILITY AND UTILIZATION IN A VIRTUAL ORGANIZATION - Preliminary Considerations IS-17
Maria Ganzha, Maciej Gawinecki, Michal Szymczak, Grzegorz Frackowiak, Marcin Paprzycki, Myon-Woong Park, Yo-Sub Han and Y. T. Sohn

(MULTI-)AGENT SYSTEMS TECHNOLOGY AND E-COMMERCE IS-27
Rainer Unland

S-CUBE: ENABLING THE NEXT GENERATION OF SOFTWARE SERVICES IS-29
Klaus Pohl

INTERNET TECHNOLOGY

FULL PAPERS

COLLABORATIVE OLAP WITH TAG CLOUDS - Web 2.0 OLAP Formalism and Experimental Evaluation 5
Kamel Aouiche, Daniel Lemire and Robert Godin

BSBC: TOWARDS A SUCCINCT DATA FORMAT FOR XML STREAMS 13
Stefan Böttcher, Rita Hartel and Christian Heinzemann

IMPLEMENTATION OF A NEW SCHEDULING POLICY IN WEB SERVERS 22
Ahmad S. Al Sa'deh and Adnan H. Yahya

BRINGING TOGETHER WHAT TOGETHER BELONGS - Applying Web Services to Couple SOA and Grid in Smaller Environments 30
Carsten Kleiner and Arne Koschel

DYNAMIC SLA NEGOTIATION BASED ON WS-AGREEMENT 38
Antoine Pichot, Oliver Wäldrich, Wolfgang Ziegler and Philipp Wieder

BLACKBIRD MONITORING SYSTEM - Performance Analysis and Monitoring in Information Systems 46
João P. Germano, Alberto R. Silva and Fernando M. Silva

OFF-THE-RECORD SECURE CHAT ROOM 54
Jiang Bian, Remzi Seker, Umit Topaloglu and Coskun Bayrak

DEVELOPING OPEN TRAVEL ALLIANCE-BASED ONTOLOGY OF GOLF
Agnieszka Cieslik, Maria Ganzha and Marcin Paprzycki 62

SHORT PAPERS

EVALUATION OF A READ-OPTIMIZED DATABASE FOR DYNAMIC WEB APPLICATIONS
Anderson Supriano, Gustavo M. D. Vieira and Luiz E. Buzato 73

ANALYSIS, DESIGN AND IMPLEMENTATION OF IDS USING DATA MINING
B. V. Patel and B. B. Meshram 81

A CONSTRAINT-AWARE QUERY OPTIMIZER FOR WEB-BASED DATA INTEGRATION
Jing Lu and Bernhard Mitschang 87

A TUPLE SPACE WEB SERVICE FOR DISTRIBUTED PROGRAMMING - Simplifying Distributed Web Services Applications
George C. Wells, Barbara Mueller and Loïc Schulé 93

A DESCRIPTIVE APPROACH FOR THE LIFECYCLE SUPPORT OF DISTRIBUTED WEB-BASED SYSTEMS
Frederic Majer, Martin Nussbaumer and Martin Gaedke 101

WEB SERVICE COMPOSITION USING THE WEB SERVICES MANAGEMENT LAYER
Niels Joncheere, Bart Verheecke, Viviane Jonckers, Sofie van Hoecke, Gregory van Seghbroeck and Bart Dhoedt 109

USING ONTOLOGIES TO IMPROVE PERFORMANCE IN A WEB SYSTEM - A Web Caching System Case of Study
Carlos Guerrero, Carlos Juiz and Ramon Puigjaner 117

A BROADCASTING ALGORITHM USING ADJUSTABLE TRANSMISSION RANGES IN MOBILE AD HOC NETWORKS
Toshihiko Sasama, Yasuhiro Abe and Hiroshi Masuyama 123

TOWARDS MKDA: A DATA MINING SEMANTIC WEB SERVICE
Vincenzo Cannella, Giuseppe Russo and Roberto Pirrone 129

A SURVEY ON WEB SERVICE DISCOVERING AND COMPOSITION
Elena del Val Noguera and Miguel Rebollo Pedruelo 135

A MACHINE LEARNING APPROACH WITH VERIFICATION OF PREDICTIONS AND ASSISTED SUPERVISION FOR A RULE-BASED NETWORK INTRUSION DETECTION SYSTEM
José Ignacio Fernández-Villamor and Mercedes Garijo 143

FORENSIC CHARACTERISTICS OF PHISHING - Petty Theft or Organized Crime?
Stephen McCombie, Paul Watters, Alex Ng and Brett Watson 149

A LOGIC PROGRAMMING MODEL FOR WEB RESOURCES
Giulio Piancastelli and Andrea Omicini 158

AN ANALYSIS OF RELATIONAL STORAGE STRATEGIES FOR PARTIALLY STRUCTURED XML
Yasser Abdel Kader, Barry Eaglestone and Siobhán North 165

A NEW CONCEPT FOR REAL-TIME WEB GAMES - Developing Highly Real-Time Web Games
Yoshihiro Kawano, Masahiro Miyata, Dai Hanawa and Tatsuhiro Yonekura 171

DQRDFS - Towards a Semantic Web Enhanced with Data Quality
Ismael Caballero, Eugenio Verbo, Coral Calero and Mario Piattini 178

A MATHEMATICAL FORMULATION OF A MODEL FOR LANDFORM ATTRIBUTES REPRESENTATION FOR APPLICATION IN DISTRIBUTED SYSTEMS
Leacir Nogueira Bastos, Rossini Pena Abrantes and Brauliro Gonçalves Leal 184

AN EFFICIENT STREAMING ALGORITHM FOR EVALUATING XPATH QUERIES
Yangjun Chen 190

INTERNET ACCESS QUALITY MONITOR
Bruno P. Ramos, Vasco N. G. J. Soares and Alexandre J. P. D. Fonte 197

RAPID VIRTUAL DESIGN AND SYSTEM DEVELOPMENT BASED ON EXTENDED MVC-BASED WEB APPLICATION FRAMEWORK AND INTERACTIVE XML PRODUCT MODEL
Cao Yan and Yang Lina 202

POSTERS

A CONCURRENCY CONTROL MODEL FOR MULTIPARTY BUSINESS PROCESSES
Juha Puustjärvi 209

MODELING THE WEB AS A FOREST OF TREES
Fathi Tenzakhti 216

GEO-GAMING: THE MOBILE MONOPOLY EXPERIENCE
Mao Li, M. J. O'Grady and G. M. P. O'Hare 220

NDT & METRICA V3 - An Approach for Public Organizations based on Model Driven Engineering
M. J. Escalona, J. J. Gutiérrez, J. A. Ortega, I. Ramos and G. Aragón 224

USING CONTENT SYNDICATION TECHNOLOGIES IN DISTRIBUTING AND PUBLISHING INFORMATION TO REACH ALL USERS
Serena Pastore 228

WORKFLOWS IN CONTENT MANAGEMENT SYSTEMS
Pedro Pico and Alberto Rodrigues da Silva 232

EVALUATION OF K-/LATTICE-CLUSTERING ALGORITHMS FOR RANDOM WIRELESS MULTI-HOP NETWORKS
Toshihiko Sasama, Ryo Monde and Hiroshi Masuyama 236

HARMONY - A FRAMEWORK FOR AUTOMATIC WEB SERVICE COMPOSITION
Viorica R. Chifu, Ioan Salomie, Emil Ş t. Chifu and Constantin Pârţac 240

A SECURE WEB APPLICATION PROVIDING PUBLIC ACCESS TO HIGH-PERFORMANCE DATA INTENSIVE SCIENTIFIC RESOURCES - ScalaBLAST Web Application
Darren Curtis, Elena Peterson and Christopher Oehmen 244

MDA-BASED DEVELOPMENT OF DATA-DRIVEN WEB APPLICATIONS
Attila Adamkó and Lajos Kollár 252

A DEVELOPMENT INFRASTRUCTURE FOR WEB SERVICES
Dionisis X. Adamopoulos 256

STRUCTURING DESIGN ACTIVITIES IN OPEN PROGRAMMABLE NETWORKS
Dionisis X. Adamopoulos 260

A WEB-BASED SYSTEM TO REDUCE THE NOSOCOMIAL INFECTION IMPACT IN HEALTCARE UNITS
Hugo Rigor, José Machado, António Abelha, José Neves and Carlos Alberto 264

TRANSACTION SUPPORT FOR INTERACTIVE WEB APPLICATIONS
David Paul, Mark Wallis, Frans Henskens and Michael Hannaford 269

XML-IS: ONTOLOGY-BASED INTEGRATION ARCHITECTURE
Christophe Cruz and Christophe Nicolle 273

IMPLEMENTING CONTENT SHARING AND SESSION HAND-OFF BETWEEN WEB BROWSERS - An Integration of SIP Stack into Mozilla Firefox Web Browser
Michael O. Adeyeye and Neco Ventura 278

DECENTRALIZED DIAGNOSIS FOR BPEL WEB SERVICES
Lina Ye and Philippe Dague 283

MULTIAGENT DESIGN FOR DYNAMIC JOB-SHOP SCHEDULING USING PASSI
Claudio Cubillos, Silvana Roncagliolo and Leonardo Espinoza 288

TOWARDS AN AGENT FRAMEWORK FOR A PASSENGER TRANSPORTATION VIRTUAL ENTERPRISE
Claudio Cubillos and Daniel Cabrera 292

XML DATA INTEGRATION IN PEER-TO-PEER DATA MANAGEMENT SYSTEMS
Tadeusz Pankowski 296

TOWARDS EFFICIENT CRYPTOGRAPHY FOR PRIVACY PRESERVING DATA MINING IN DISTRIBUTED SYSTEMS
Emmanouil Magkos and Vassilis Chrissikopoulos 301

E-LEARNING

FULL PAPERS

HAPTICS AND EXTENSIBLE 3D IN WEB-BASED ENVIRONMENTS FOR E-LEARNING AND SIMULATION
Felix G. Hamza-Lup and Ivan Sopin 309

A GENERAL FRAMEWORK FOR REPLICATED EXPERIMENTS IN VIRTUAL 3D ENVIRONMENTS
D. Biella and W. Luther 316

CASE STUDIES VIA THE WEB FOR CONTINUOUS PROFESSIONAL DEVELOPMENT - Use of the ViCoCITY Web-based Case Study Support Tool
James A. Redmond, Audrey Stenson and Alan Mullally 324

TRANSFORMING A COMPETENCY MODEL TO ASSESSMENT ITEMS
Onjira Sitthisak, Lester Gilbert and Hugh C. Davis 333

FORMALIZING A MODEL TO REPRESENT AND VISUALIZE CONCEPT SPACES IN E-LEARNING ENVIRONMENTS
Antonina Dattolo and Flaminia L. Luccio 339

SHORT PAPERS

E-LEARNING ACTIVITIES DESIGN AND INDIVIDUAL LEARNING STYLES - Case Study
Cláudia Fernandes and Luís Rocha 349

E-LEARNING AS A SOLUTION TO THE TRAINING PROBLEMS OF SMEs - A Multiple Case Study
Andrée Roy and Louis Raymond 356

LEARNING PERSONALIZATION - Design Solutions in an e-Learning System
Ileana Trandafir, Ana-Maria Borozan and Alexandru Balog 364

TOWARDS WEB 2.0 DRIVEN LEARNING ENVIRONMENTS
Mohamed Amine Chatti, Daniel Dah, Matthias Jarke and Gottfried Vossen 370

A TRANSLITERATION ENGINE FOR ASIAN LANGUAGES
Sathiamoorthy Manoharan 376

LAUNCHING AN E-LEARNING SYSTEM IN A SCHOOL - Cross-European e-/m-Learning System UNITE: A Case Study
Maja Ćukušić, Andrina Granić and Ivan Maršić 380

E-LEARNING TOOLS FOR WOUND IMAGE UNDERSTANDING
Augustin Prodan, Mădălina Rusu, Remus Câmpean and Rodica Prodan 388

USING EVALUATION AS A QUALITY ASSURANCE TOOL IN THE DEVELOPMENT OF SERIOUS GAMES - A Case Study based on the PRIME Game
Jannicke Baalsrud Hauge, Heiko Duin and Manuel Oliveira 394

OPEN SOURCE LMS CUSTOMIZATION - A Moodle Stadistical Control Aplication
Carlos Muñoz, Miguel Ángel Conde, Jorge Reyero and Francisco José García 402

DESIGNING 3D COLLABORATIVE VIRTUAL ENVIRONMENTS TO UTILIZE THE PEDAGOGICAL BENEFITS OF CSCL
Th. Tsiatsos and A. Konstantinidis 408

TEACHING PROGRAMMING WITH A COMPETITIVE ATTITUDE TO FOSTER GROUP SPIRIT
Pedro Guerreiro and Katerina Georgouli 414

IMS-CLD: A NEW SPECIFICATION FOR LEARNING SCENARIOS IN COPES
Azeddine Chikh, Lamia Berkani and Akila Sarirete 422

USING ALTERNATE REALITY GAMES TO SUPPORT THE TEACHING OF MODERN FOREIGN LANGUAGES
Thomas M. Connolly 428

PREDISPOSITION-BASED INTELLIGENT TUTORING SYSTEM - Adaptive User Profiling in Human-Computer Interaction
Andrzej Niesler and Gracja Wydmuch 435

POSTERS

DESIGN OF DIGITAL EDUCATIONAL MATERIALS FOR PRIMARY EDUCATION
Isabel Cuadrado Gordillo and Inmaculada Fernández Antelo 443

INTRODUCING E-LEARNING 2.0 IN SME - A Practical Guide
Ileana Hamburg 448

APPLICATION OF A WEB-BASED EDUCATION SYSTEM IN INDUSTRIAL PROCESSES
Perfecto Mariño, Miguel Ángel Domínguez, Santiago Otero and Miguel Merino 452

INTERACTIVE, COLLABORATIVE AND ADAPTATIVE LEARNING TOOLS - The TexMat Example
A. M. Breda, E. M. Rocha and M. M. Rodrigues 456

USEFUL E-LEARNING PROCESS DESCRIPTIONS
Steffen Mencke, Fritz Zbrog and Reiner Dumke 460

PROACTIVE AUTONOMOUS RESOURCE ENRICHMENT FOR E-LEARNING
Steffen Mencke, Dmytro Rud, Fritz Zbrog and Reiner Dumke 464

INTEROPERABILITY GUIDELINES FOR DIGITAL LIBRARY OF EDUCATIONAL RESOURCES AND SERVICES
Kurilovas Eugenijus and Kubilinskienė Svetlana 468

ABOUT THE BENEFITS OF EXPLOITING NATURAL LANGUAGE PROCESSING TECHNIQUES FOR E-LEARNING
Diana Pérez-Marín, Ismael Pascual-Nieto and Pilar Rodríguez 472

AUTHOR INDEX 477

DQRDFS
Towards a Semantic Web Enhanced with Data Quality

Ismael Caballero, Eugenio Verbo
Indra Software Labs, UCLM - Indra Research and Development Institute
Ronda de Toledo s/n – 13004 Ciudad Real, Spain
{icaballerom, emverbo}@indra.es

Coral Calero, Mario Piattini
ALARCOS Research Group, Information Systems and Technology Departament
UCLM - Indra Research and Development Institute, Paseo de la Universidad 4 s/n – 13071 Ciudad Real, Spain
{Coral.Calero, Mario.Piattini}@uclm.es

Keywords: Data Quality, Semantic Web, DQRDFS, DQMetadata, ISO 15939, Data Quality Measurement, Quantity of Data Quality.

Abstract: Nowadays, data is of critical importance as a resource. Using data of poor quality can be the source of several problems when developing a project. The World Wide Web is currently the main showcase for a vast amount of data. It would be desirable that machines can process the quality of the data contained on the Web Documents. This paper introduces a new view of the Semantic Web based on the concept of Quantity of Data Quality (QDQ), in which Data Quality issues will be used as a basis to enable machines to process the Semantic Web Documents for different activities like information retrieval or document filtering. This view can open new challenges in Semantic Webs oriented to improve users' satisfaction with the Internet.

1 INTRODUCTION

Advances in software technologies and networks have lead to the development of new Information Systems running on the Web, as a means for organizations to be as close as possible to all their customers, stakeholders and other organizations. This kind of system allows companies to publish, via the web, documents containing data related to the tasks which must be carried out by specific users. If data is not of high enough quality, then users cannot correctly complete their projects. So it is important to take care with data quality (hereafter DQ) in order to ensure that users achieve a better standard of project. Up to now, users have been responsible for assessing the quality of the data, since they are the only ones who are able to understand it and its meaning.

In their definition of Semantic Web, (Bernes-Lee et al., 2001) state that "*the Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation*". This definition opens the possibility for machines to understand data contained in web documents: if they can understand it, they can also manage the quality of it. Furthermore, machines could apply this managerial capability to doing tasks like discriminating documents attending to their data quality levels before deploying them to users.

But in order to turn these proposals into reality, some value must be added to the Semantic Web, and to the way in which Semantic Web Documents can be performed.

This paper has a twofold goal: (1) to provide readers with a brief background of DQ (section 2) and (2) to show how we have applied DQ fundamentals in order to enhance the quality of Web Documents for Semantic Web (section 3).

2 DATA QUALITY BACKGROUND

2.1 Data Quality Dimensions

Data is said to be of quality if it fits the purpose for which is used (Strong et al., 1997). One of the most

interesting strategies for tackling the study of DQ for a context, is to break it down into "minor qualities" known as **DQ dimensions** (Lee et al., 2006). Each scenario requires some dimensions which best fit the use of the data. The sets of DQ Dimensions usable in a context are known as **DQ model**. Literature shows several examples of DQ models for specific scenarios: medical and healthcare (Al-Hakim, 2007), decision support system (Gendron and D'Onofrio, 2002), or web (Caro et al., 2007), to name a few. ISO, at this moment, is working on the draft of the ISO/IEC 25012 standard (ISO-25012, 2007), a part of the SQUARE family that will propose a DQ model for IS. In any case, the generic classification proposed by (Strong et al., 1997) has been widely used. These authors group DQ dimensions into four categories making reference to the point of view from which DQ can be observed: **Intrinsic DQ** (dimensions of accuracy, objectivity, credibility, reputation) refers to the quality of the data itself; **Accessibility DQ** category contains dimensions (accessibility, Access security) providing meaning about the extent to which data can be accessed; **Contextual DQ** (Relevancy, Value-Added, Timeliness, Completeness, Amount of data) refers to those DQ dimensions which deal with the use of specific data in a context; **Representational DQ** category (interpretability, ease of understanding, concise representation, consistent representation) is centred on those characteristics of the representation of data which make it usable. A more complete definition of the meaning of these dimensions can be found through DQ literature, the most interesting works being those proposed by (Batini and Scannapieco, 2006, Lee et al., 2006, English, 1999).

2.2 Measuring and Assessing DQ

Let us give the name *stakeholder* to any person or process involved in the use of the data or of resources which have data. Any stakeholder will need to assess how good a piece of data is for the task to be executed. We would like to highlight the difference between the concepts of "measuring DQ" and "assessing the DQ level" of a piece of data for a task. For each DQ Dimension belonging to the DQ model used for the assessment, some measurements must be taken.

Both measurement and assessment are going to depend on the intended use of the data and on the nature of the DQ dimension (which determine the measurement method (ISO/IEC, 2000)). For measuring, a base or a derived measure must be drawn. In this case, a measurement method or a measurement function is required. On the other hand, it could possibly be said that for the assessment and indicator might be enunciated.

According to literature, typical derived DQ measures have a measurement function based on the percentage of the Number of Data Units which do or do not satisfy a DQ criterion (Batini and Scannapieco, 2006, Lee et al., 2006). This fact confers to the measurement of a ratio scale (see Figure 1):

$$DQ_{Measure} = 1 - \frac{NumberOfDataUnitsNotSatisfyingADQCriterion}{TotalNumberOfDataUnits}$$

Figure 1: Typical DQ Measure.

In the formula of Figure 1, there are two base measures: *NumberOfDataUnitsNotSatisfyingACriterion* and *TotalNumberOfDataUnits*. The measurement method for both consists of counting a number of data units. For the second one, there is only one problem, which is counting all data units of the piece of data. In the first, the counting is limited to those data units affected by the criterion. A criterion is usually defined as a business rule to warranty the soundness of the data (English, 1999, Loshin, 2001, Wang, 1998). The result of deciding if the data unit satisfies the criterion can be "*True*" or "*False*". So, in order to obtain a value for the measure, a count of data units having obtained a "*true*" value must be done. But the intrinsic difficulty is addressed at defining how the data unit satisfies the criterion. Sometimes, some metadata is necessary for each piece of data to complete its meaning in order to be able to decide whether or not it satisfies the criteria. To make a decision, a rule based on this metadata is needed. This rule can consist of objectively or subjectively determining if the value of metadata belongs to a given domain. (Naumann and Rolker, 2000) identify the following as possible sources for values of metadata: a stakeholder, the information manufacturing process or even the same data store. Different authors in the DQ field agree that values for metadata coming from users are probably subjective, whereas the ones coming from the proper data stores are objective.

Having to add some metadata to the data, a new problem arises: how to attach the metadata to the data and how to store it conveniently. In (Wang et al., 1995), a possible solution for the relational model is proposed. It consists of tagging data: attach the DQ metadata as if it were another common attribute. It could be seen as a way of semantic annotation. (Caballero et al., 2007) propose another solution based on (Wang et al., 1995) for XML.

They propose an XML Schema named DQXSD that allows making such annotations for XML files (see figure 2): The **qualityData** is used as the root of the XML document; **Entity** is anything containing data (a relational schema, an XML documented). Each entity can have **attributes,** the DQ of which must be studied, like relational attributes or elements in XML files. Authors use the name **measurableConcept** for DQ Dimensions in order to align their model to ISO/IEC 15939 (ISO/IEC, 2000). Finally, for each *measurableConcept*, zero or more **DQMetadata-Attributes** are defined, and given a value which is used to assess the DQ level of each entity regarding the measurable concept.

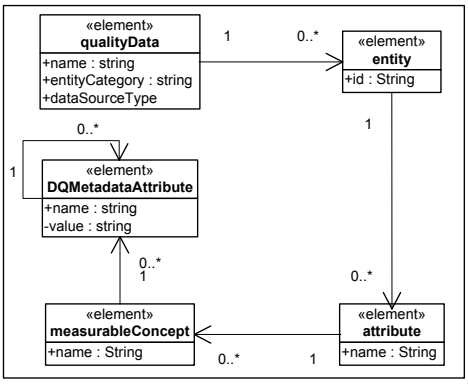

Figure 2: DQXSD (Caballero et al., 2007).

Next, we are going to explain how these fundamentals could be introduced into the Semantic Web.

3 DQRDFS: DQ AT SEMANTIC WEB

The main aim of this paper is to enable a new perspective in which machines can automatically process the quality of the data contained in the Semantic Web Documents in order to increase user's satisfaction with Semantic Web applications in tasks like information retrieval or semantic searches. This implies that machines need to measure and assess this data by making corresponding DQ semantic annotations in "*such way that can be used for more effective discovery, automation, integration and reuse across various applications* (Guha et al., 2003)"

For this reason, we are going to show how to integrate DQ issues into Semantic Web processing by following (Wang et al., 1995)'s ideas through the proposal of (Caballero et al., 2007), but adapted to RDF. In the "*traditional*" view, data on Semantic Web is modelled like a ***directed labelled graph***, wherein each node corresponds to a resource (subjects and objects) and each arc is labelled with a predicate. As a first approach to integrate DQ issues in Semantic Web, what we propose in this paper is to annotate RDF with values (metadata following DQ nomenclature or *DQMetadataAttribute* following (Caballero et al., 2007)) corresponding to DQ dimensions (*measurableConcepts*) which are of interest for the different *stakeholders*. This metadata might have an objective value used to compute a measure. Having measures for all best fitting DQ Dimensions, machines can process an assessment for the Semantic Web Document. This assessment represents the perception of the DQ for a Web Document of a stakeholder for a given application. We have named it as ***Quantity of DQ*** **(QDQ).**

The QDQ could be interpreted as a weight. It enables viewing the Semantic Web as a ***weighted directed graph*** for a specific task and stakeholder. This view will open new fields in machine-processing data having as basis DQ: for instance, Semantic Searchers can delimit the quantity of found results, showing to the users only those whose QDQ is within an acceptance threshold range (not only the "relevance"); or ordering the results according to a ranking model (Ding et al., 2005) based on DQ requirements. Another kind of application that can be improved is that oriented to automate a task, like the one described for (Bernes-Lee et al., 2001).

In order to achieve this goal, several challenges must be tackled: (1) identify which attributes must be studied from DQ point of view, (2) how to identify which are the proper *measurableConcepts* for those resources and how to identify the necessary metadata (when required) to make the measurements, (3) how to get values for that metadata and how to annotate them, and finally (4) how to compute the QDQ according to the perception of DQ of different groups of stakeholders through their selected dimensions.

3.1 Identify Attributes to be Annotated

The first step in order to enable DQ in Semantic Web consists of identifying from the Data Quality User Requirement Specification (DQ-URS) which elements need to be studied. The elements are related to the level of granularity at which the study is necessary. (Ding et al., 2005) identify the following levels of granularity according to the levels in which Semantic Web can be queried: RDF Database, Semantic Web Document (SWD), RDF subgraph or Semantic Web. According to DQXSD by (Caballero et al., 2007), our entities are going to

be the **RDF.** Their attributes are *Subject, Predicates, Object* and *Sentences.* In this approach we are going to focus only on **sentences,** our aim being to write DQ Sentences about sentences. This process is known as **reification** (Daconta et al., 2003).

For instance, let us consider the sentences "*Buddy owns business"* and "*business has-Website http://www.buddy.com"*-borrowed directly from (Daconta et al., 2003) and showed in figure 3-. Let's imagine that somebody could be interested in knowing, for instance, how reliable, reputable or timely the sentences are.

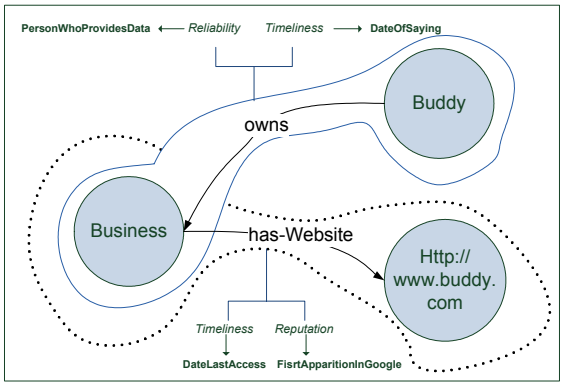

Figure 3: DQ Dimensions for different sentences.

3.2 Identify DQ Dimensions and Related Metadata

Once the attributes susceptible to study have been identified, the next step is to identify the DQ dimensions or measurable concepts that best fit the problem of assessing DQ for these elements. The easiest way to tackle this problem is to choose as a guide a suitable specific DQ Model (see section 2.1)

As previously mentioned, and according to the nature of each DQ Dimension, some metadata would be required. Sometimes, metadata can already be part of the RDF file or it may be necessary to add it. For instance, in Figure 2, a stakeholder could need two DQ dimensions (*Timeliness* and *Reliability*) to compute the QDQ of the sentence "*Buddy owns a Business*". Some metadata complementing the meaning of the *Reliability* dimension is required. Let us suppose that it has been decided that knowing the *PersonWhoProvidesData* can help to interpret and determine if the sentence is worthy or not. Please, note that on one hand we have the values corresponding to metadata for measuring a DQ Dimension by using a measurement method, and on the other hand we have the measures of the DQ Dimension used to calculate the QDQ for the predicate by aggregating those values through an indicator.

3.3 Getting and Annotating Values for Metadata

This is the great challenge since it implies three key aspects in measuring DQ:

(1) *Who must provide these values?*
(2) *How and where to store these values?* and
(3) *How to get a representative value for different values of the same* ***DQMetadataAttribute*** *for all stakeholders in order to calculate the QDQ?*

The main response to question (1) might be found in social annotations like those in Web 2.0 (e.g. del.icio.us or flickr) (Bao et al., 2007). This situation enables on one hand, the possibility of easily getting values for the same metadata, with all the connotations and backgrounds of each user. And on the other hand, it is possible to determine through users' experiences the relationships amongst the most important DQ Dimensions when assessing the DQ from their corresponding backgrounds in order to create specific DQ Models for each context.

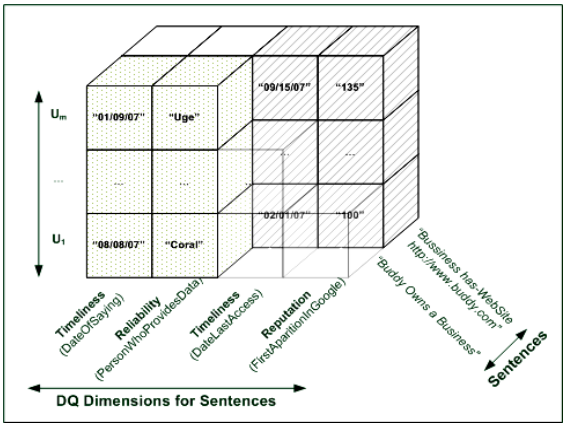

Figure 4: Conceptual Representation of DQ metadata.

To answer question (2), in Figure 4, a cube is shown with the following information: On the X axis, we have the DQ dimensions; the Y axis shows all users having annotated a value; the Z axis gathers all possible sentences of an RDF. Each individual block stores the value V_{ijk} for a metadata given for a DQ dimension D_i by a user U_j for the sentence S_k. So, column *i* contains values of metadata for the DQ dimension D_i, whereas the row *j* gathers the values given for a user u_j for all DQ dimensions involved in the evaluation of a sentence S_k. Since not all dimensions are implied at the same time in the calculus of QDQ for each sentence, not all values are required. Each block of figure 4 has been particularized with values for the example proposed in Figure 3. All these values must be stored together with the RDF in a RDF Server or in a XML

Database. The DQXSD proposed by (Caballero et al., 2007) is used for describing how to attach and store the value for metadata to their corresponding sentence (an attribute for DQXSD).

In order to make this DQXSD operative in this context, we have developed a counterpart RDF Schema, which is shown in Figure 5. This Schema is what we have named **DQRDFS.** As an example of its use, figure 6 shows how the sentence "*Buddy owns a Business"* and its corresponding *measurableConcepts* (see figure 3) can be represented by using the proposed DQRDFS.

```
<?xml version='1.0' encoding='UTF-8'?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:dq="'http://alarcos.inf-cr.uclm.es/dqrdf/1.0'"
 xmlns:rdfs="'http://www.w3.org/2000/01/rdf-schema ">
<rdfs:Class rdf:about="&dq;DQMetadataAttribute"
 rdfs:label="DQMetadataAttribute">
<rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
</rdfs:Class>
<rdf:Property rdf:about="&dq;DQMetadataAttributes"
 rdfs:label="DQMetadataAttributes">
<rdfs:domain rdf:resource="&dq;measurableConcept"/>
<rdfs:range rdf:resource="&rdfs;Class"/>
</rdf:Property>
<rdfs:Class rdf:about="&dq;attribute"
 rdfs:label="attribute">
<rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
</rdfs:Class>
<rdf:Property rdf:about="&dq;attributes"
 rdfs:label="attributes">
<rdfs:domain rdf:resource="&dq;entity"/>
<rdfs:range rdf:resource="&rdfs;Class"/>
</rdf:Property>
<rdfs:Class rdf:about="&dq;entity"
 rdfs:label="entity">
<rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
</rdfs:Class>
<rdf:Property rdf:about="&dq;id"
 rdfs:label="id">
<rdfs:domain rdf:resource="&dq;entity"/>
<rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>
<rdfs:Class rdf:about="&dq;measurableConcept"
 rdfs:label="measurableConcept">
<rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
</rdfs:Class>
<rdf:Property rdf:about="&dq;measurableConcepts"
 rdfs:label="measurableConcepts">
<rdfs:domain rdf:resource="&dq;attribute"/>
<rdfs:range rdf:resource="&rdfs;Class"/>
</rdf:Property>
<rdf:Property rdf:about="&dq;name"
 rdfs:label="name">
<rdfs:domain rdf:resource="&dq;DQMetadataAttribute"/>
<rdfs:domain rdf:resource="&dq;attribute"/>
<rdfs:domain rdf:resource="&dq;measurableConcept"/>
<rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>
<rdf:Property rdf:about="&dq;nestedEntity"
 rdfs:label="nestedEntity">
<rdfs:domain rdf:resource="&dq;DQMetadataAttribute"/>
<rdfs:range rdf:resource="&rdfs;Class"/>
</rdf:Property>
<rdf:Property rdf:about="&dq;value"
 rdfs:label="value">
<rdfs:domain rdf:resource="&dq;DQMetadataAttribute"/>
<rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>
</rdf:RDF>
```

Figure 5: DQRDFS: a RDF Schema supporting DQ.

For the last question, it is important to realise that having several values for each metadata, it is necessary to give a global representative value for all provided values. If given values are numbers, an example of this representing global value could be an arithmetic mean; should these values for metadata be subjective linguistic labels, then an aggregation method is required, like the one proposed by (Herrera-Viedma et al., 2006).

```
<?xml version="1.0"?>
<rdf:RDF [...]
 xmlns:dq="http://alarcos.inf-cr.uclm.es/ontologies/dqmo#" >
<rdf:Description>
 <earl:asserts rdf:parteType='Resource'>
  <rdf:subject>
   <rdf:Description rdf:about="#Buddy"></rdf:subject>
  <rdf:predicate>
   <RDFNSId:owns><RDFNSId='#business'>
  </rdf:predicate>
 </earl:asserts>
 <dq:entity rdf:about="&dq;entity_Instance"
  dq:id="entity1"
  rdfs:label="entity_Instance">
  <dq:attributes rdf:resource="&dq;sentence_Instance"/>
 </dq:entity>
 <dq:attribute rdf:about="&dq;sentence_Instance"
  <dq:name="sentence1" rdfs:label="sentence_Instance">
   <dq:measurableConcepts
    df:resource="&dq;measurableConcept_Instance_1"/>
   <dq:measurableConcepts
    rdf:resource="&dq;measurableConcept_Instance_2"/>
 </dq:attribute>
 <dq:measurableConcept
rdf:about="&dq;measurableConcept_Instance_1"
  dq:name="Reliability"
  rdfs:label="measurableConcept_Instance_1">
  <dq:DQMetadataAttributes
   rdf:resource="&dq;metadataAtt_Instance_1"/>
 </dq:measurableConcept>
 <dq:measurableConcept
rdf:about="&dq;measurableConcept_Instance_2"
  dq:name="Timeliness"
  rdfs:label="measurableConcept_Instance_2">
  <dq:DQMetadataAttributes
   rdf:resource="&dq;metadataAtt_Instance_2"/>
 </dq:measurableConcept>
  <dq:DQMetadataAttribute
rdf:about="&dq;metadataAtt_Instance_1"
  dq:name="PersonWhoProvidesData"
  rdfs:label="metadataAtt_Instance_1">
  <dq:value> "Uge"</dq:value>
  <dq:value> "Coral"</dq:value>
 </dq:DQMetadataAttribute>
 <dq:DQMetadataAttribute
rdf:about="&dq;metadataAtt_Instance_2"
  dq:name="DateOfSaying"
  rdfs:label="metadataAtt_Instance_2">
  <dq:value> "01/09/07"</dq:value>
  <dq:value> "08-08-2007"</dq:value>
 </dq:DQMetadataAttribute>
</rdf:Description>
</rdf:RDF>
```

Figure 6: An example of DQRDF.

3.4 Calculating QDQ

Once representative values for each measurable concept have been obtained, the next step is to calculate the QDQ for each sentence. As can be seen

in figures 3 and 4, for each QDQ a set of DQ Dimensions is involved. The value of QDQ, as previously said, must be calculated by aggregating the corresponding measures for the required *measurable Concepts* as an Indicator (ISO/IEC, 2000), taking into account the relationships between the different DQ Dimensions. An interesting proposal for calculating the QDQ, which takes into account the possible relationship between DQ Dimensions, is the one by (Caro et al., 2007), in which a Bayesian Network (BN) is implemented for their own DQ model for calculating the level of Representational DQ of Educational Web Portals.

4 CONCLUSIONS

This paper has introduced some fundamentals of DQ and has highlighted the importance of annotating DQ issues of Semantic Web in order to have an improved web through QDQ Concept. This QDQ enables a view of the Web as a weighted directed graph which would open new challenges in machine-processing Semantic Web Documents in order to optimize users' satisfaction. In the future we will deal with refining and validating the DQRDFS. A study of how to extend the proposal to the remainder of the elements of RDF is also planned.

ACKNOWLEDGEMENTS

This research is part of the projects ESFINGE (TIN2006-15175-C05-05) supported by the Spanish Ministerio of Educación y Ciencia and MECENAS (PBI06-0024) supported by the Consejería de Educación y Ciencia of Junta de Comunidades de Castilla – La Mancha.

REFERENCES

Al-Hakim, L. (2007). Procedure for Mapping Information Flow: A case of Surgery Management Process. In *Information Quality Management: Theory and Applications* (Ed, Al-Hakim, L.) Idea Group Publishing, Hershey, PA, USA, pp. 168-188.

Bao, S., Wu, X., Fei, B., Xue, G., Su, Z. and Yu, Y. (2007) Optimizing Web Search Using Social Annotations. In *WWW'05 Proceedings*. ACM Press, Banff, Alberta, Canada, pp. 501-510.

Batini, C. and Scannapieco, M. (2006) *Data Quality: Concepts, Methodologies and Techniques*, Springer-Verlag Berlin Heidelberg, Berlin.

Bernes-Lee, T., Hendler, J. and Lassila, O. (2001) The Semantic Web. *Scientific American.*

Caballero, I., Verbo, E. M., Calero, C. and Piattini, M. (2007) A Data Quality Measurement Information Model based on ISO/IEC 15939 *In 12th ICIQ Proceedings*. MIT, Cambridge, MA. Pp. 393-408.

Caro, A., Calero, C. and Piattini, M. (2007) Development process of the operational version of PDQM *In The 8th WISE Proceedings*. Springer-Berlag, Nancy, France.

Daconta, M., Obsrt, L. and Smith, K. (2003) *The Semantic Web: A guide to the future of XML. Web Services and Knowledge Managemen*t, Willey Inc, Indianápolis, Indiana.

Ding, L., Finin, T., Joshi, A., Peng, Y., Pan, R. and Reddivari, P. (2005) Search on the Semantic Web. *IEEE Computer*, 38, 62-69.

English, L. (1999) *Improving Data Warehouse and Business Information Quality: Methods for reducing costs and increasing Profits*, Willey & Sons, New York, NY, USA.

Gendron, M. and D'Onofrio, M. J. (2002) Formulation of a Decision Support Model Using Quality Attributes In *7th ICIQ. MIT*, Cambridge, MA, USA, pp. 305-316.

Guha, R., McCool, R. and Miller, E. (2003) Semantic search. *In Proceedings of the 12th WWW*. ACM Press, Budapest, Hungary, pp. 700-709.

Herrera-Viedma, E., Pasi, G. and Lopez-Herrera, A. (2006) Evaluating the Information Quality of Web Sites: A Quality Methodology Based on Fuzzy Computing with Words. *JASIST*, 54, 538-549.

ISO-25012 (2007) ISO/IEC 25012: Software Engineering - Software Quality Requirements and Evaluation (SQuaRE) - Data Quality Model (Draft). Vol. 2007 ISO/IEC.

ISO/IEC (2000) ISO/IEC 15939. *Information Technology - Software Measurement Process*.

Lee, Y. W., Pipino, L. L., Funk, J. D. and Wang, R. Y. (2006) *Journey to Data Quality*, Massachussets Institute of Technology, Cambridge, MA, USA.

Loshin, D. (2001) Data Quality and Business Rules. In Information and Database Quality (Eds, Piattini, M., Calero, C. and Genero, M.) Kluwer Academic Publishers.

Naumann, F. and Rolker, C. (2000) Assessment Methods for Information Quality Criteria *In Fifth ICIQ Proceedings*. MIT, Cambridge, MA, USA, pp. 148-162.

Strong, D. M., Lee, Y. W. and Wang, R. Y. (1997) Data Quality in Context. *Commun. ACM*,, 40, 103-110.

Wang, R. Y. (1998) A Product Perspective on Total Data Quality Management. *Commun. ACM*, 41, 58-65.

Wang, R. Y., Reddy, M. and Kon, H. (1995) Towards quality data: An attribute-based approach. *Decis. Supp. Syst.*, 13, 349-372.