

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

University of Dortmund, Germany

Madhu Sudan

Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Stefano Spaccapietra Jeff Z. Pan
Philippe Thiran Terry Halpin
Steffen Staab Vojtech Svatek
Pavel Shvaiko John Roddick (Eds.)

Journal on Data Semantics XI

Volume Editors

Stefano Spaccapietra
EPFL-IC-IIF-LBD, Lausanne, Switzerland
E-mail: stefano.spaccapietra@epfl.ch

Jeff Z. Pan
The University of Aberdeen, UK; jeff.z.pan@abdn.ac.uk

Philippe Thiran
Namur University, Belgium; philippe.thiran@fundp.ac.be

Terry Halpin
Neumont University, Salt Lake City, UT, USA; terry@neumont.edu

Steffen Staab
Universität Koblenz-Landau, Germany; staab@uni-koblenz.de

Vojtech Svatek
University of Economics, Prague, Czech Republic; svatek@vse.cz

Pavel Shvaiko
TasLab, University of Trento, Povo, Italy; pavel@dit.unitn.it

John Roddick
Flinders University, Adelaide, Australia; roddick@cs.flinders.edu.au

Library of Congress Control Number: Applied for

CR Subject Classification (1998): H.2, H.3, I.3, C.2, H.4.3, D.3

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/
Web and HCI

ISSN	0302-9743 (Lecture Notes in Computer Science)
ISSN	1861-2032 (Journal on Data Semantics)
ISBN-10	3-540-92147-8 Springer Berlin Heidelberg New York
ISBN-13	978-3-540-92147-9 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© Springer-Verlag Berlin Heidelberg 2008
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12585670 06/3180 5 4 3 2 1 0

The LNCS Journal on Data Semantics

Computerized information handling has changed its focus from centralized data management systems to decentralized data exchange facilities. Modern distribution channels, such as high-speed Internet networks and wireless communication infrastructure, provide reliable technical support for data distribution and data access, materializing the new, popular idea that data may be available to anybody, anywhere, anytime. However, providing huge amounts of data on request often turns into a counterproductive service, making the data useless because of poor relevance or inappropriate level of detail. Semantic knowledge is the essential missing piece that allows the delivery of information that matches user requirements. Semantic agreement, in particular, is essential to meaningful data exchange.

Semantic issues have long been open issues in data and knowledge management. However, the boom in semantically poor technologies, such as the Web and XML, has boosted renewed interest in semantics. Conferences on the Semantic Web, for instance, attract big crowds of participants, while ontologies on their own have become a hot and popular topic in the database and artificial intelligence communities.

Springer's *LNCS Journal on Data Semantics* aims at providing a highly visible dissemination channel for remarkable work that in one way or another addresses research and development on issues related to the semantics of data. The target domain ranges from theories supporting the formal definition of semantic content to innovative domain-specific application of semantic knowledge. This publication channel should be of the highest interest to researchers and advanced practitioners working on the Semantic Web, interoperability, mobile information services, data warehousing, knowledge representation and reasoning, conceptual database modeling, ontologies, and artificial intelligence.

Topics of relevance to this journal include:

- Semantic interoperability, semantic mediators
- Ontologies
- Ontology, schema and data integration, reconciliation and alignment
- Multiple representations, alternative representations
- Knowledge representation and reasoning
- Conceptualization and representation
- Multimodel and multiparadigm approaches
- Mappings, transformations, reverse engineering
- Metadata
- Conceptual data modeling
- Integrity description and handling
- Evolution and change
- Web semantics and semi-structured data
- Semantic caching

- Data warehousing and semantic data mining
- Spatial, temporal, multimedia and multimodal semantics
- Semantics in data visualization
- Semantic services for mobile users
- Supporting tools
- Applications of semantic-driven approaches

These topics are to be understood as specifically related to semantic issues. Contributions submitted to the journal and dealing with semantics of data will be considered even if they are not from the topics in the list.

While the physical appearance of the journal issues is like the books from the well-known Springer LNCS series, the mode of operation is that of a journal. Contributions can be freely submitted by authors and are reviewed by the Editorial Board. Contributions may also be invited, and nevertheless carefully reviewed, as in the case for issues that contain extended versions of the best papers from major conferences addressing data semantics issues. Special issues, focusing on a specific topic, are coordinated by guest editors once the proposal for a special issue is accepted by the Editorial Board. Finally, it is also possible that a journal issue be devoted to a single text.

The journal published its first volume in 2003 (LNCS 2800). That initial volume, as well as volumes II (LNCS 3360), V (LNCS 3870), VIII (LNCS 4380), IX (LNCS 4601), and this volume XI represent the annual occurrence of a special issue devoted to publication of selected extended versions of the best conference papers from conferences of the year before. Volumes III and VI were special issues on a dedicated topic. Volume III (LNCS 3534), coordinated by guest editor Esteban Zimányi, addressed Semantic-Based Geographical Information Systems, while volume VI (LNCS 4090), coordinated by guest editors Karl Aberer and Philippe Cudre-Mauroux, addressed Emergent Semantics. Volumes IV (LNCS 3730) and VII (LNCS 4244) were “normal” volumes, built from spontaneous submissions on any of the topics of interest to the journal.

The Editorial Board comprises an Editor-in-Chief (with overall responsibility), a Co-editor-in-Chief, and several members. The Editor-in-Chief has a four-year mandate. Members of the board have a three-year mandate. Mandates are renewable and new members may be elected anytime.

We are happy to welcome you to our readership and authorship, and hope we will share this privileged contact for a long time.

Stefano Spaccapietra
Editor-in-Chief
<http://lbdwww.epfl.ch/e/Springer/>

JoDS Volume XI

To foster the dissemination of the best ideas and results, the *Journal on Data Semantics* (JoDS) pursues a policy that includes annually publishing extended versions of the best papers from selected conferences whose scope encompasses or intersects the scope of the journal.

This initiative is motivated by the difference in goals between conferences and journals. Conferences usually have a faster turnaround and a focused audience, but they have to enforce space limitation and a fixed time frame, with no chances for improving a paper by producing multiple versions. In contrast, journals offer more space, room for debate and refinement, and are usually considered the real archival venue.

Therefore, the publication of an extended version of a conference paper is a much appreciated opportunity for researchers to widely disseminate a significantly improved presentation of their work, where they can develop the appropriate motivations, reasoning, results and comparative analysis. Moreover, by gathering the best papers from various conferences, JoDS special issues provide a unique opportunity for researchers to find in a single publication every year the best of ongoing research in the field of data semantics.

For this issue, papers from the following 2006 international workshops and conferences were invited:

- The Second International Workshop on Semantic Web-Enabled Software Engineering (SWESE 2006), held at the 5th International Semantic Web Conference (ISWC 2006). November 6, 2006, Athens, GA, USA. Guest editor: Jeff Z. Pan
- InterDB 2006, The Second International Workshop on Database Interoperability, Held in conjunction with ICDE 2006. April 3, 2006, Atlanta, USA. Guest editor: Philippe Thiran
- The Second International Workshop on Object-Role Modeling (ORM 2006), Montpellier, France, November 2–3, 2006, in conjunction with the On The Move Federated Conferences (OTM 2006). Guest editor: Terry Halpin
- EKAW 2006 - 15th International Conference on Knowledge Engineering and Knowledge Management - Managing Knowledge in a World of Networks, Podebrady, Czech Republic, October 2006. Guest Editors: Steffen Staab and Vojtech Svatek
- The International Workshop on Ontology Matching (OM-2006), co-located with the 5th International Semantic Web Conference (ISWC 2006), November 5, 2006, Athens, Georgia, USA. Guest editor: Pavel Shvaiko
- Third International Workshop on Conceptual Modeling for Geographic Information Systems (CoMoGIS 2006), and First International Workshop on Semantic Web Applications: Theory and Practice (SemWAT 2006), both in conjunction with the 25th International Conference on Conceptual Modeling (ER 2006). Guest editor: John Roddick

Papers were invited based on their quality, relevance and significance, and the viability of extending their results. Extended versions prepared by authors were subject to the traditional two-round scholarly review process, and the authors were required to respond to all concerns expressed by the reviewers before papers were accepted. Eight papers were eventually accepted for publication in this issue.

The selection of SWESE best papers eventually resulted in the acceptance of two papers.

The first paper “Experiences in the Design of Semantic Services Using Web Engineering Methods and Tools,” by Brambilla, Ceri, Celino, Cerizza, Della Valle, Facca, Turati, and Tzviskou, shows how classical software engineering methods (such as formal business process development and automatic code generation) combine with semantic methods and tools (i.e., ontology engineering, semantic service annotation and discovery) to forge a new approach to software development for the Semantic Web. In the paper, the authors present their experience in the participation to the Semantic Web Service Challenge 2006, where the proposed approach achieved very good results in solving the proposed problems.

The second paper “Automatically Generated Model Transformations Using Ontology Engineering Space,” by Roser and Bauer, presents an approach to using the semantic technologies to improve cross-organizational modeling by automated generation of model transformations. By automated generation of mappings it offers new possibilities for the integration of domain specific languages and ‘legacy’ models in a plug&play manner, making it easier for new organizations to join collaborations.

The extended version of the InterDB 2006 paper “The Harmony Integration Workbench” by Mork, Seligman, Rosenthal, Korb and Wolf presents a workbench for schema integration in which multiple tools share a common knowledge repository. The tool supports the interoperation of existing tools through a common view of schemas and mappings. The workbench is intended to help integration engineers to use and to select the best tool for an integration problem.

One extended article was selected from ORM 2006. The paper “Using ORM-Based Models as a Foundation for a Data Quality Firewall in an Advanced Generation Data Warehouse (Extended version)” by Piprani reports on the use of the Object-Role Modeling (ORM) approach to establish a data quality firewall architecture for an industrial data warehouse. The paper describes a metadata repository framework for extracting data from heterogeneous sources, cleansing and transforming the data for loading into the data warehouse, and then generating specific data marts. Techniques for performing detailed quality checks are described, along with auditing and control measures. The outlined approach has proven its value in a variety of industrial applications, typically resulting in 35-40% savings in overall project costs.

The EKAW 2006 conference contributed with an extended version of the paper “Discovering Semantic Sibling Groups from Web Documents with XTREEM-SG” by Brunzel and Spiliopoulou. The paper addresses a novel issue in ontology learning: the exploitation of term collections embedded in HTML mark-up structures. The particular targets are semantic siblings, i.e., co-hyponyms and co-meronyms in the taxonomic structure of an ontology. The advantage of this approach, compared to NLP-based ontology learning approaches, is relative independence on linguistic resources. Experimental evaluation of the sibling mining algorithm was made

against a gold-standard data set, and several variations of input setting were systematically tested.

OM-2006 contributed its best paper, namely, “Exploring the Semantic Web as Background Knowledge for Ontology Matching” by Sabou, d'Aquin, and Motta. This work proposes to enhance ontology matching by combining appropriate background knowledge from multiple automatically identified online ontologies. Specifically, two strategies have been devised and implemented: (i) mappings within one ontology, that is when a mapping between the candidate concepts (of the ontologies to be matched) can be discovered in a single external ontology, and (ii) cross-ontology mapping discovery, that is when a mapping is derived from two or more (external) ontologies. The approach has been evaluated on real-life data from two large agricultural thesauri, such as AGROVOC and NALT.

The selection from the workshops held in conjunction with the ER 2006 conference resulted in two extended papers being accepted for JoDS: "Time-Aggregated Graphs for Modeling Spatio-Temporal Networks" by George and Shekhar, and "An Architecture for Emergent Semantics" by Herschel, Heese, Bleiholder and Czekay.

The first paper proposes a novel method for storing and efficiently querying spatio-temporal data. A detailed logical and physical structure is presented along with an analytical analysis of the storage requirements of the proposed model compared to other approaches. Following this a new algorithm is presented for finding the shortest path using the proposed data model. The algorithm is then proven correct and its worst-case runtime is formally analyzed.

The second paper introduces a universal architecture for emergent semantics using a central repository within a multi-user environment, based on linguistic theories. Based on this architecture, an implementation of an information retrieval system supporting term queries on standard information retrieval corpora is shown, which incorporates feedback on the retrieval results directly into the actual document representations, thus improving future retrievals.

SWESE 2006 Co-chair

Jeff Z. Pan

University of Aberdeen, UK

InterDB 2006 Co-chair

Philippe Thiran

Namur University, Belgium

ORM 2006 Co-chair

Terry Halpin

Neumont University, Salt Lake City,
Utah, USA

EKA W 2006 Co-chairs

Steffen Staab
Vojtech Svatek

University of Koblenz-Landau, Germany
University of Economics, Prague, Czech Republic

OM-2006 Co-chair

Pavel Shvaiko

TasLab, Informatica Trentina, Italy*

ER Workshops 2006 Co-chair

John Roddick

Flinders University, Australia

Reviewers

We would like to express our gratitude to the following colleagues who helped in the review process by contributing detailed reviews of the submitted papers and provided invaluable feedback to the authors and editors:

Colin Atkinson	University of Mannheim, Germany
Nathalie Aussenac-Gilles	IRIT- CNRS Toulouse, France
Ken Baclawski	Northeastern University, USA
Philipp Cimiano	University of Karlsruhe, Germany
Philippe Cudre-Mauroux	EPFL, Switzerland
Denise de Vries	Flinders University, Australia
Ken Evans	ORM Foundation, UK
Dragan Gasevic	Simon Fraser University Surrey, Canada
Marc-Philippe Huget	University of Savoy, France
Michael N. Huhns	University of South Carolina, USA
Zakaria Maamar	Zayed University, UAE
David Martin	SRI International, USA
Sjir Nijssen	PNA Group, The Netherlands
Natasha Noy	Stanford University, USA
Christine Parent	University of Lausanne, Switzerland
Adrian Paschke	Technical University Dresden, Germany
Allison L. Powell	Corporation for National Research Initiative, USA
Erik Proper	Radboud University and Cap Gemini, The Netherlands
Gerald Reif	University of Zurich, Switzerland
Simone Santini	Universidad Autonoma de Madrid, Spain
Pavel Smrz	Technical University of Brno, Czech Republic
Heiner Stuckenschmidt	University of Mannheim, Germany
Michael Twidale	University of Illinois at Champaign, USA

* The work on this JoDS issue was performed in part while at the University of Trento, Italy.

Chad A. Williams
Huimin Zhao

University of Illinois at Chicago, USA
University of Wisconsin-Milwaukee, USA

JoDS Editorial Board

Co-editors-in-Chief

Lois Delcambre
Stefano Spaccapietra

Portland State University, USA
EPFL, Switzerland

Members of the Board

Carlo Batini	Università di Milano Bicocca, Italy
Alex Borgida	Rutgers University, USA
Shawn Bowers	University of California Davis, USA
Tiziana Catarci	Università di Roma La Sapienza, Italy
David W. Embley	Brigham Young University, USA
Jerome Euzenat	INRIA Alpes, France
Dieter Fensel	University of Innsbruck, Austria
Nicola Guarino	National Research Council, Italy
Jean-Luc Hainaut	FUNDP Namur, Belgium
Ian Horrocks	University of Manchester, UK
Arantza Illarramendi	Universidad del País Vasco, Spain
Larry Kerschberg	George Mason University, USA
Michael Kifer	State University of New York at Stony Brook, USA
Tok Wang Ling	National University of Singapore, Singapore
Shamkant B. Navathe	Georgia Institute of Technology, USA
Antoni Olivé	Universitat Politècnica de Catalunya, Spain
José Palazzo M. de Oliveira	Universidade Federal do Rio Grande do Sul, Brazil
Christine Parent	Université de Lausanne, Switzerland
Klaus-Dieter Schewe	Massey University, New Zealand
Heiner Stuckenschmidt	University of Mannheim, Germany
Katsumi Tanaka	University of Kyoto, Japan
Yair Wand	University of British Columbia, Canada
Eric Yu	University of Toronto, Canada
Esteban Zimányi	Université Libre de Bruxelles (ULB), Belgium

Table of Contents

Second SWESE Workshop

Experiences in the Design of Semantic Services Using Web Engineering Methods and Tools	1
<i>Marco Brambilla, Stefano Ceri, Irene Celino, Dario Cerizza, Emanuele Della Valle, Federico M. Facca, Andrea Turati, and Christina Tziviskou</i>	
Automatic Generation and Evolution of Model Transformations Using Ontology Engineering Space	32
<i>Stephan Roser and Bernhard Bauer</i>	

Second International Workshop on Database Interoperability (InterDB 2006)

The Harmony Integration Workbench	65
<i>Peter Mork, Len Seligman, Arnon Rosenthal, Joel Korb, and Chris Wolf</i>	

International Workshop on Object-Role Modeling (ORM 2006)

Using ORM-Based Models as a Foundation for a Data Quality Firewall in an Advanced Generation Data Warehouse (Extended Version)	94
<i>Baba Piprani</i>	

15th International Conference on Knowledge Engineering and Knowledge Management – Managing Knowledge in a World of Networks (EKAW 2006)

Discovering Groups of Sibling Terms from Web Documents with XTREEM-SG	126
<i>Marko Brunzel and Myra Spiliopoulou</i>	

International Workshop on Ontology Matching (OM-2006)

Exploring the Semantic Web as Background Knowledge for Ontology Matching	156
<i>Marta Sabou, Mathieu d'Aquin, and Enrico Motta</i>	

ER 2006 Workshops

Time-Aggregated Graphs for Modeling Spatio-temporal Networks 191
 Betsy George and Shashi Shekhar

An Architecture for Emergent Semantics 213
 Sven Herschel, Ralf Heese, Jens Bleiholder, and Christian Czekay

Author Index 235