

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

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

New York University, NY, USA

Doug Tygar

University of California, Berkeley, CA, USA

Moshe Y. Vardi

Rice University, Houston, TX, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Mokrane Bouzeghoub Carole Goble
Vipul Kashyap Stefano Spaccapietra (Eds.)

Semantics of a Networked World

Semantics for Grid Databases

First International IFIP Conference, ICSNW 2004
Paris, France, June 17-19, 2004
Revised Selected Papers



Springer

Volume Editors

Mokrane Bouzeghoub
CNRS - Université de Versailles, Laboratoire PRiSM
45, avenue des Etats-Unis, 78035 Versailles, France
E-mail: mok@prism.uvsq.fr

Carole Goble
University of Manchester, Department of Computer Science
Oxford Road, Manchester, M13 9PL, UK
E-mail: carole@cs.man.ac.uk

Vipul Kashyap
Clinical Informatics R&D, Partners HealthCare System, Inc.
93 Worcester St., P.O. Box 81905, Wellesley, MA 02481, USA
E-mail: vkashyap1@partners.org

Stefano Spaccapietra
Swiss Federal Institute of Technology Lausanne, EPFL-IC-LBD
1015 Lausanne, Switzerland
E-mail: stefano.spaccapietra@epfl.ch

Library of Congress Control Number: 2004113830

CR Subject Classification (1998): H.2, H.3, H.4, C.2.4, D.2.12, D.4.3-4

ISSN 0302-9743

ISBN 3-540-23609-0 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 is a part of Springer Science+Business Media
springeronline.com

© 2004 International Federation for Information Processing, Hofstrasse 3, A-2361 Laxenburg, Austria
Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Protago-TeX-Production GmbH
Printed on acid-free paper SPIN: 11339694 06/3142 5 4 3 2 1 0

In Memory of Professor Yahiko Kambayashi

Prof. Kambayashi was for many years an active member of the IFIP 2.6 Working Group, which organized the conference whose proceedings make up this volume.

Prof. Kambayashi suddenly passed away in February 2004. He leaves us with the memory of a wonderfully competent, energetic and kind colleague.

The members of the Working Group are very sad to lose one of their best friends. They wish to offer this volume to his memory as a testimony of their esteem for the late Prof. Kambayashi.

S. Kambayashi

Y. Kambayashi

~~Y. Kambayashi~~

K. Kambayashi

Y. Kambayashi

Y. Kambayashi

S. Kambayashi

A. Kambayashi

~~Y. Kambayashi~~

Y. Kambayashi

Y. Kambayashi

Y. Kambayashi

Y. Kambayashi

Y. Kambayashi

Y. Kambayashi

Y. Kambayashi

Y. Kambayashi

Y. Kambayashi

Preface

The explosion in data exchange fostered by the success of the Web has restated semantics as a kernel issue in the development of services providing data and information to users and applications worldwide. This newly designated conference series on “Semantics for the Networked World” unifies into a single framework the previous series on “Database Semantics” and “Visual Database Systems” that the IFIP WG 2.6 has been offering since 1985. Whereas the intent of the conference series is to explore interesting research issues related to semantics, the theme for the 2004 edition is “Semantics for Grid Databases”. Grid computing, a new field concentrating on “flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources (also referred to as virtual organizations)”, has gathered momentum in the context of providing shared infrastructures for large-scale scientific computations and data analysis. Similarly, P2P computing has attracted substantial attention.

Currently, attention is devoted to the provision of middleware services to make computational resources interoperable at the technical level and to increase the efficiency of use of physical resources. However, as Grid and P2P computing infrastructures are being increasingly adopted, they are likely to have typical problems of information overload that manifest themselves in any large-scale infrastructure for information and application sharing (e.g., the WWW). The need for resource discovery, application and service interoperability, integration and composition manifest themselves in these infrastructures. The ability to interoperate at the semantic level will largely determine the continued success and utilization of these infrastructures.

This working conference focused on issues of semantic interoperability of the information and services provided and manipulated by Grid and P2P computing systems. The purpose of the conference, as for its predecessors, was to provide an active forum for researchers and practitioners for presentation and exchange of research results and practical management of databases, this time applied to large-scale systems used for managing and sharing scientific information. The technical program of the conference, which was held during June 17–19, 2004 in Paris, represented the state of the art in developing principles and practices related to the management of semantics in large-scale networked (Grid and P2P) information infrastructures. It featured two invited talks, by Norman Paton (Databases and the Grid: JDBC in WSDL, or Something Altogether Different?) and Karl Aberer (Emergent Semantics Systems), a tutorial by Dave Berry (The State of the Grid), and a panel on “What Can and Should We Do for the Grid?”. The rate of acceptance of the research track was 1 out of 3, and the papers were presented in the following sessions: integration, peer-to-peer computing, semantics for scientific applications, interoperability and mediation, and global services and schemas. This was followed by a posters program that sought to showcase research ideas that are in various stages of progress and show promise.

Finally, the conference chairs would like to thank all the people who worked hard to bring this conference into being. This includes the members of the Program Committee, the invited and tutorial speakers and the authors of the various submissions who put forth their research ideas. Last, but not least, we would like to thank the local organizers in Paris — Daniela Grigori, Stéphane Lopes and Annick Baffert — whose efforts made this conference feasible and operational. We hope the attendees had stimulating and interesting exchanges of innovative ideas, possibly with some meaningful advances in this new and emerging area of research.

June 2004

Mokrane Bouzeghoub
Carole Goble
Vipul Kashyap
Stefano Spaccapietra

Organization

General Chair

Mokrane Bouzeghoub, University of Versailles, France

Program Co-chairs

Carole Goble, University of Manchester, UK

Stefano Spaccapietra, Federal Institute of Technology, Lausanne, Switzerland

Vipul Kashyap, National Library of Medicine, Bethesda, MD, USA

Steering Committee

Tiziana Catarci, University “La Sapienza”, Rome, Italy

Karl Aberer, Swiss Federal Institute of Technology, Lausanne, Switzerland

Marc Scholl, University of Konstanz, Germany

Local Arrangements

Daniela Grigori, University of Versailles, France

Stéphane Lopes, University of Versailles, France

Program Committee Members

Karl Aberer, EPFL Lausanne, Switzerland

Mario Cannataro, University Magna Græcia, Catanzaro, Italy

Isabel Cruz, University of Illinois at Chicago, USA

Vassilis Christophides, ICS-FORTH, Crete, Greece

Christine Collet, Institut National Polytechnique de Grenoble, France

Nigel Collier, National Institute of Informatics, Tokyo, Japan

Isabel F. Cruz, University of Illinois at Chicago, USA

Ewa Deelman, University of Southern California, USA

Peter Fankhauser, IPSI, Darmstadt, Germany

Ian Foster, Argonne National Laboratory, USA

Alon Halevy, University of Washington, USA

Maurizio Lenzerini, University “La Sapienza”, Rome, Italy

Ling Liu, Georgia Institute of Technology, USA

Bertram Ludaescher, San Diego Supercomputing Center, USA

Robert Meersman, Free University of Brussels (VUB), Belgium

Reagan Moore, San Diego Supercomputer Center (SDSC), USA

James Myers, Pacific Northwest National Laboratory, USA

John Mylopolous, University of Toronto, Ontario, Canada

Aris Ouksel, University of Illinois at Chicago, USA

Norman Paton, University of Manchester, UK

Dave Pearson, Oracle, UK

Munindar Singh, North Carolina State University, Raleigh, USA

Steffen Staab, Karlsruhe University, Germany

Paul Watson, University of Newcastle upon Tyne, UK

Hai Zhuge, Chinese Academy of Sciences, China

Esteban Zimanyi, Free University of Brussels (ULB), Belgium

Table of Contents

Invited Talks

Databases and the Grid: JDBC in WSDL, or Something Altogether Different?	1
<i>Norman W. Paton</i>	

Emergent Semantics Systems	14
<i>Karl Aberer, Tiziana Catarci, Philippe Cudré-Mauroux, Tharam Dillon, Stephan Grimm, Mohand-Said Hacid, Arantza Illarramendi, Mustafa Jarrar, Vipul Kashyap, Massimo Mecella, Eduardo Mena, Erich J. Neuhold, Aris M. Ouksel, Thomas Risse, Monica Scannapieco, Fèlix Saltor, Luca de Santis, Stefano Spaccapietra, Steffen Staab, Rudi Studer, Olga De Troyer</i>	

Integration

User-Specific Semantic Integration of Heterogeneous Data: The SIRUP Approach	44
<i>Patrick Ziegler, Klaus R. Dittrich</i>	

Towards a Mediation System Framework for Transparent Access to Largely Distributed Sources. The MediaGrid Project	65
<i>Christine Collet, Khalid Belhajjame, Gilles Bernot, Christophe Bobineau, Gennaro Bruno, Beatrice Finance, Fabrice Jouanot, Zoubida Kedad, David Laurent, Fariza Tahi, Genoveva Vargas-Solar, Tuyet-Trinh Vu, Xiaohui Xue</i>	

Null Values Revisited in Prospect of Data Integration	79
<i>Guy de Tré, Rita de Caluwe, Henri Prade</i>	

Peer to Peer

Semantic-Based Query Routing and Heterogeneous Data Integration in Peer-to-Peer Semantic Link Networks	91
<i>Hai Zhuge, Jie Liu, Liang Feng, Chao He</i>	

Peer Selection in Peer-to-Peer Networks with Semantic Topologies	108
<i>Peter Haase, Ronny Siebes, Frank van Harmelen</i>	

Autonomous Gossiping: A Self-Organizing Epidemic Algorithm for Selective Information Dissemination in Wireless Mobile Ad-Hoc Networks	126
<i>Anwitaman Datta, Silvia Quarteroni, Karl Aberer</i>	

Hyper: A Framework for Peer-to-Peer Data Integration on Grids	144
<i>Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, Riccardo Rosati, Guido Vetere</i>	

Semantics for Scientific Applications

Semantically Linking and Browsing Provenance Logs for E-science	158
<i>Jun Zhao, Carole Goble, Robert Stevens, Sean Bechhofer</i>	
Emergent Semantics: Towards Self-Organizing Scientific Metadata	177
<i>Bill Howe, Kuldeep Tanna, Paul Turner, David Maier</i>	
Semantic Visualization of Biochemical Databases	199
<i>Esteban Zimányi, Sabri Skhiri dit Gabouje</i>	

Interoperability and Mediation

Semantic Information Interoperability in Open Networked Systems	215
<i>Silvana Castano, Alfio Ferrara, Stefano Montanelli, Gianpaolo Racca</i>	
A New Mechanism for the Interoperability of Data Systems	231
<i>Miren I. Bagüés, Jesús Bermúdez, Arantza Illarramendi, Alberto Tablado, Alfredo Goñi</i>	
An Architecture for Recommendation Based Service Mediation	250
<i>Bhaskar Mehta, Claudia Niederée, Avare Stewart, Claudio Muscogiuri, Erich J. Neuhold</i>	

Global Services and Schemas

Static-Dynamic Integration of External Services into Generic Business Processes	263
<i>Günter Preuner, Christian Eichinger, Michael Schrefl</i>	
Knowledge Sifter: Agent-Based Ontology-Driven Search over Heterogeneous Databases Using Semantic Web Services	278
<i>Larry Kerschberg, Mizan Chowdhury, Alberto Damiano, Hanjo Jeong, Scott Mitchell, Jingwei Si, Stephen Smith</i>	
Managing Grid Schemas Globally	296
<i>Kimio Kuramitsu</i>	

Posters

Relational Graphical Models of Computational Workflows for Data Mining	309
<i>William H. Hsu</i>	

InfoGrid: Information Resource Integration	311
<i>Nikolaos Giannadakis, Mustafa Ghanem, Y. Guo</i>	
Semantic Annotation of Classification Data for KDD Support Services . . .	313
<i>Claudia Diamantini, Domenico Potena, Maurizio Panti</i>	
Role of the Ontologies in the Context of Grid Computing and Application for the Human Disease Studies	316
<i>Maja Hadzic, Elizabeth Chang</i>	
Cluster Entries for Semantic Organization of Peer-to-Peer Network	319
<i>Nicolas Lumineau, Anne Doucet, Bruno Defude</i>	
A Peer-to-Peer Service Supporting Data Quality: Design and Implementation Issues	321
<i>Diego Milano, Monica Scannapieco, Tiziana Catarci</i>	
Analysis Patterns for the Geographic Database Conceptual Schema: An Ontology Aided Approach	323
<i>Guillermo Nudelman Hess, Cirano Iochpe, José Palazzo Moreira de Oliveira</i>	
Author Index	325