

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

TU Dortmund University, Germany

Madhu Sudan

Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbruecken, Germany

Pablo Barceló Reinhard Pichler (Eds.)

Datalog in Academia and Industry

Second International Workshop, Datalog 2.0
Vienna, Austria, September 11-13, 2012
Proceedings

Volume Editors

Pablo Barceló

Universidad de Chile, Department of Computer Science
Avenida Blanco Encalada 2120, 3er Piso, 837-0459 Santiago, Chile
E-mail: pbarcelo@dcc.uchile.cl

Reinhard Pichler

Vienna University of Technology, Institute of Information Systems
Favoritenstraße 9-11, 1040 Wien, Austria
E-mail: pichler@dbai.tuwien.ac.at

ISSN 0302-9743

e-ISSN 1611-3349

ISBN 978-3-642-32924-1

e-ISBN 978-3-642-32925-8

DOI 10.1007/978-3-642-32925-8

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012944718

CR Subject Classification (1998): H.2.4, H.4, I.2.4, C.2, H.3, H.5, D.1.6

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

© Springer-Verlag Berlin Heidelberg 2012

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.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This volume contains the proceedings of the 2012 Workshop on the Resurgence of Datalog in Academia and Industry (Datalog 2.0 2012) held during September 11–13, 2012, in Vienna. Datalog 2.0 is a workshop for Datalog researchers, implementers, and users. Its main aim is to bring everyone up to date and map out directions for future research. The first edition of this workshop was held in Oxford, UK, in March 2010. It was based on invitations only. Over the past few years, Datalog has been resurrected as a lively topic with applications in many different areas of computer science as well as industry. Owing to this renewed interest and increased level of activity, we decided to open the workshop for submissions this year.

The call for papers resulted in 17 submissions. Each paper was reviewed by at least three Program Committee members. The Program Committee accepted 12 papers, based on their technical merit and potential for stimulating discussions. We also accepted one system description and one tutorial, which show the influence of Datalog in industry and practice today.

In addition, the technical program included invited talks by Thomas Eiter (Technische Universität Wien), Yuri Gurevich (Microsoft Research), Phokion Kolaitis (University of California at Santa Cruz and IBM Research - Almaden), Oege de Moor (University of Oxford), and Marie-Laure Mugnier (University of Montpellier). It also includes invited tutorials by Todd J. Green (University of California at Davis and LogicBlox) and Axel Pollers (Siemens AG Austria). The Datalog 2.0 Workshop 2012 was colocated with two further events: the 6th International Conference on Web Reasoning and Rule Systems (RR 2012) and the 4th International Conference on Computational Models of Argument (COMMA 2012). The invited talk by Robert Kowalski (Imperial College, London) was shared by all three events. Short abstracts of two invited talks are included in the front matter of this volume. Extended abstracts are given in the main body of the proceedings.

The workshop would not have been possible without the support of many people. First of all we would like to thank Georg Gottlob, the General Chair of Datalog 2.0, for his advice and help in putting together the technical program. We are very grateful to Markus Pichlmair (Technische Universität Wien) for all his hard work in the local organization. We also acknowledge EasyChair as a great tool that has significantly simplified the whole process from receiving the submissions to producing the input for the proceedings. Finally, we would like to thank all the authors who contributed to the workshop and the Program Committee members for their effort to produce timely and wise reviews.

We hope that the success of Datalog 2.0 2012 will stimulate forthcoming versions of this workshop as well as renewed interest from the community in Datalog and its applications.

September 2012

Pablo Barceló
Reinhard Pichler

Organization

The 2012 Workshop on the Resurgence of Datalog in Academia and Industry (Datalog 2.0 2012) was organized by the Institute of Information Systems at the Vienna University of Technology.

Program Committee

General Chair

Georg Gottlob	University of Oxford, UK
---------------	--------------------------

Program Chairs

Pablo Barceló	Universidad de Chile, Chile
Reinhard Pichler	Vienna University of Technology, Austria

Members

Chitta Baral	Arizona State University, USA
Piero Bonatti	University of Naples Federico II, Italy
Loreto Bravo	Universidad de Concepción, Chile
Andrea Cali	University of London, Birkbeck College, UK
Diego Calvanese	Free University of Bozen-Bolzano, Italy
Jürgen Dix	Clausthal University of Technology, Germany
Tim Furche	University of Oxford, UK
Claudio Gutierrez	Universidad de Chile, Chile
Georg Lausen	University of Freiburg, Germany
Nicola Leone	University of Calabria, Italy
Annie Liu	State University of New York at Stony Brook, USA
Boon Thau Loo	University of Pennsylvania, USA
Raghu Ramakrishnan	Yahoo! Research, USA
Sebastian Rudolph	AIFB, Karlsruhe Institute of Technology, Germany
Tuncay Tekle	LogicBlox, USA
Mirosław Truszczyński	University of Kentucky, USA
Stijn Vansummen	Université Libre de Bruxelles, Belgium
Victor Vianu	UC San Diego, USA
Stefan Woltran	Vienna University of Technology, Austria
Peter Wood	Birkbeck, University of London, UK

Local Organization

Chair

Markus Pichlmair

Vienna University of Technology, Austria

Further Members

Ingo Feinerer

Vienna University of Technology, Austria

Reinhard Pichler

Vienna University of Technology, Austria

Additional Reviewers

Alviano, Mario

Arenas, Marcelo

Caroprese, Luciano

Kaminski, Roland

May, Wolfgang

Montali, Marco

Pieris, Andreas

Puehrer, Joerg

Ricca, Francesco

Terracina, Giorgio

Paraconsistent Modular Answer Set Programming (Abstract)

Thomas Eiter
Vienna University of Technology
`eiter@kr.tuwien.ac.at`

Paraconsistent reasoning is a well-studied approach to deal with inconsistency in logical theories in a way such that inference does not explode. It has specifically been considered in the area of knowledge representation and reasoning for a range of different formalisms, including also non-monotonic formalisms such as logic programming. In the last years, there has been increasing interest in datalog-based formalisms, including traditional Answer Set Programming, and extensions of the formalisms to encompass modularity, possibly in a distributed environment, have been conceived.

In this talk, we shall address the issue of paraconsistency for modular logic programs in a datalog setting, under the answer set semantics for logic programs. The two orthogonal aspects of modularity and paraconsistency for this semantics, which is subsumed by Equilibrium Logic, may be approached on different grounds. We shall consider developments on these aspects, including proposals for modularity and paraconsistency of answer set programs developed at TU Wien. For the latter particular emphasis is given to the issue of incoherence, i.e., non-existence of answer sets due to the lack of stability caused by cyclic dependencies of an atom from its default negation. We shall then consider possible combinations of the two aspects in a single formalism. In the course of this, we shall discuss issues and challenges regarding semantics and evaluation, both in theory and for practical concerns.

This work is a joint effort with Minh Dao-Tran, Michael Fink, Thomas Krennwallner and others, and supported by the project P20841 “Modular HEX-Programs” of the Austrian Science Fund (FWF).

A Retrospective on Datalog 1.0 (Abstract)

Phokion G. Kolaitis

University of California Santa Cruz & IBM Research - Almaden

`kolaitis@cs.ucsc.edu`

Datalog was introduced in the early 1980s as a database query language that embodies a recursion mechanism on relational databases and, thus, overcomes some of the inherent limitations in the expressive power of relational algebra and relational calculus. In the ensuing decade, Datalog became the subject of an in-depth investigation by researchers in database theory. This study spanned a wide spectrum of topics, including query processing and optimization of Datalog programs, the delineation of the expressive power and computational complexity of Datalog queries, and the exploration of the semantics and the expressive power of extensions of Datalog with comparison operators and negation. The investigation of these topics entailed extensive interaction of database theory with finite model theory and, in particular, with finite-variable logics and pebble games suitable for analyzing the expressive power of Datalog and its variants. From the early 1990s on, there has been a fruitful and far-reaching interaction between Datalog and constraint satisfaction; this interaction, which continues today, has contributed to the understanding of tractable cases of the constraint satisfaction problem, but has also given rise to new results about the expressive power and the computational complexity of Datalog queries.

The aim of this talk is to reflect on some of the aforementioned topics, highlight selected results, and speculate on future uses and applications of Datalog.

Table of Contents

Abstracts

Paraconsistent Modular Answer Set Programming	IX
<i>Thomas Eiter</i>	
A Retrospective on Datalog 1.0	XI
<i>Phokion G. Kolaitis</i>	

Invited Presentations

LogicBlox, Platform and Language: A Tutorial	1
<i>Todd J. Green, Molham Aref, and Grigoris Karvounarakis</i>	
Datalog: A Perspective and the Potential	9
<i>Yuri Gurevich</i>	
Existential Rules: A Graph-Based View (Extended Abstract)	21
<i>Marie-Laure Mugnier</i>	
How (Well) Do Datalog, SPARQL and RIF Interplay?	27
<i>Axel Polleres</i>	

Regular Papers, System Descriptions and Tutorials

Magic-Sets for <i>Datalog</i> with Existential Quantifiers	31
<i>Mario Alviano, Nicola Leone, Marco Manna, Giorgio Terracina, and Pierfrancesco Veltri</i>	
On the CRON Conjecture	44
<i>Tom J. Ameloot and Jan Van den Bussche</i>	
Order in Datalog with Applications to Declarative Output	56
<i>Stefan Brass</i>	
A Broad Class of First-Order Rewritable Tuple-Generating Dependencies	68
<i>Cristina Civili and Riccardo Rosati</i>	
Datalog Development Tools (Extended Abstract)	81
<i>Onofrio Febbraio, Giovanni Grasso, Nicola Leone, Kristian Reale, and Francesco Ricca</i>	
Query Rewriting Using Datalog for Duplicate Resolution	86
<i>Jaffer Gardezi and Leopoldo Bertossi</i>	

Reasoning about Knowledge in Distributed Systems Using Datalog	99
<i>Matteo Interlandi</i>	
Declarative Datalog Debugging for Mere Mortals	111
<i>Sven Köhler, Bertram Ludäscher, and Yannis Smaragdakis</i>	
Inconsistency-Tolerant Query Rewriting for Linear Datalog+/-	123
<i>Thomas Lukasiewicz, Maria Vanina Martinez, and Gerardo I. Simari</i>	
Confluence Analysis for Distributed Programs: A Model-Theoretic Approach	135
<i>William R. Marczak, Peter Alvaro, Neil Conway, Joseph M. Hellerstein, and David Maier</i>	
Business Network Reconstruction Using Datalog	148
<i>Daniel Ritter and Till Westmann</i>	
Data Exchange in Datalog Is Mainly a Matter of Choice	153
<i>Domenico Saccà and Edoardo Serra</i>	
Optimizing Large-Scale Semi-Naïve Datalog Evaluation in Hadoop	165
<i>Marianne Shaw, Paraschos Koutris, Bill Howe, and Dan Suciu</i>	
Logical Foundations of Continuous Query Languages for Data Streams	177
<i>Carlo Zaniolo</i>	
Author Index	191