Lecture Notes in Computer Science 2967

Edited by G. Goos, J. Hartmanis, and J. van Leeuwen

Springer
Berlin
Heidelberg
New York
Hong Kong
London
Milan
Paris
Tokyo

Sergey Melnik

Generic Model Management

Concepts and Algorithms

Springer

Series Editors

Gerhard Goos, Karlsruhe University, Germany
Juris Hartmanis, Cornell University, NY, USA
Jan van Leeuwen, Utrecht University, The Netherlands

Author

Sergey Melnik
Microsoft Corporation
One Microsoft Way, Redmond, WA 98052-6399, USA
E-mail: melnik@microsoft.com

Library of Congress Control Number: 2004104636

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

ISSN 0302-9743
ISBN 3-540-21980-3 Springer-Verlag Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Protago-TeX-Production GmbH
Printed on acid-free paper SPIN: 11007593 06/3142 5 4 3 2 1 0

Preface

Many challenging problems facing information systems engineering involve the manipulation of complex metadata artifacts, or *models*, such as database schemas, interface specifications, or object diagrams, and *mappings* between models. The applications that solve metadata manipulation problems are complex and hard to build. The goal of generic model management is to reduce the amount of programming needed to develop such applications by providing a database infrastructure in which a set of high-level algebraic operators, such as Match, Merge, and Compose, are applied to models and mappings as a whole rather than to their individual building blocks.

This dissertation presents an initial study of the concepts and algorithms for generic model management. We describe the first prototype of a generic model management system, introduce the algebraic operators that are used to manipulate models and mappings, clarify the semantics of the operators, and develop novel algorithms for implementing them. In particular, we present an innovative algorithm based on fixpoint computation that is used for implementing the generic operator Match, which finds correspondences between two models. Using the prototype and the operators presented in the dissertation, we develop solutions for several practically relevant problems, such as change propagation and reintegration.

April 2004 Sergey Melnik

Acknowledgements

I would like to express my deep gratitude to everyone who helped me shape the ideas explored in this dissertation, either by giving technical advice or encouraging and supporting my work in many other ways.

I enjoyed a rare privilege of collaborating closely with several distinguished database researchers. This dissertation would not have come into existence without their hands-on advice and motivation.

Prof. Erhard Rahm supervised and guided my work from the very first day. He gave me the opportunity to conduct this doctoral research and helped me make the right strategic decisions at many forks along the way. He kept me on track while allowing me to broaden my research horizon in tangential areas. His insightful comments, which densely filled the margins of each draft that I gave to him, gave rise to many creative ideas.

Prof. Hector Garcia-Molina invited me to Stanford University and taught me the art of turning hard research challenges into fun and expressing my thoughts clearly using examples. From him I learned that solid research requires patience: for example, he suggested that a draft of our joint paper [Melnik, Garcia-Molina, Rahm 2002] needed more polishing and so we missed a conference deadline. Later that paper, which underpins Part III of the dissertation, received the Best Student Paper Award at the Intl. Conf. on Data Engineering.

Prof. Emeritus Gio Wiederhold showed me what it takes to step back and see a big picture, and yet keep the details in focus. He gave me the opportunity to collaborate in the DARPA DAML project at Stanford and to get a foretaste of metadata management problems in the context of interoperation on the Semantic Web.

Dr. Philip A. Bernstein has been the driving force behind the emerging research area of generic model management, the subject of the dissertation. His vision papers and talks inspired much of the work done in this thesis. His insightful suggestions on our joint papers and his guidance in designing the first prototype for model management, which is presented in Part I of the dissertation, were invaluable.

Prof. Alon Halevy helped me keep my spirits high while I worked on Part II, a more theoretical part of the dissertation. His encouragement and advice made this work a real pleasure.

I am grateful to Profs. Serge Abiteboul, Paolo Atzeni, Stefano Ceri, Martin Kersten, Renée Miller, and Gerhard Weikum for helpful discussions.

I would like to thank my colleagues and friends in the Database Groups in Leipzig and Stanford, the members of the Graduate Programme on Knowledge Representation in Leipzig, and the colleagues in the RDF Core Working Group at the World-Wide Web Consortium for fruitful exchanges of ideas. The members of the Stanford Database Group helped me conduct the user study presented in Part III.

I am indebted to Drs. Stefan Decker, Andreas Paepcke, Bertram Ludäscher, Felix Naumann, and Arturo Crespo for their support and many informal discussions, which helped me put my academic research into perspective.

I owe very special thanks to my wife Teresa. Her love and energy constantly recharged my forces. She has been my perpetual source of creativity and inspiration, in so many respects.

This dissertation is dedicated to my parents, Tanja and Juri, who are truly the origin of all great things that ever happened to me.

The contributions of the above people made the work on this dissertation a rewarding and memorable experience. I thank you all.

April 2004 Sergey Melnik

Table of Contents

List of Figures

List of Tables