

Proceedings

Sixteenth Working Conference on

Reverse Engineering

13-16 October 2009
Lille, France

Sponsored by
Reengineering Forum

Technical co-sponsor
IEEE Computer Society Technical Council on Software Engineering (TCSE)

IEEE
 computer
society

 IEEE

Edited by
Andy Zaidman
Giuliano Antoniol
Stéphane Ducasse

Working Conference on **Reverse Engineering 2009**



Published by the IEEE Computer Society
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314

IEEE Computer Society Order Number P3867
BMS Part Number: CFP09090-PRT
ISSN Number 1095-1350
ISBN 978-0-7695-3867-9



Proceedings

**16th Working Conference
on Reverse Engineering (WCRE 2009)**

Proceedings

16th Working Conference on Reverse Engineering (WCRE 2009)

13th – 16th October 2009 – Lille, France

Edited by

Andy Zaidman, Giuliano Antoniol and Stéphane Ducasee

Sponsored by



Reengineering Forum

Technical Co-sponsor



IEEE-CS Technical Council on
Software Engineering

In cooperation with



INRIA Lille-Nord Europe



SOCGER LAB
Software Cost-effective Change
and Evolution Research Lab



Los Alamitos, California
Washington • Tokyo



All rights reserved.

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries may photocopy beyond the limits of US copyright law, for private use of patrons, those articles in this volume that carry a code at the bottom of the first page, provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Other copying, reprint, or republication requests should be addressed to: IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, P.O. Box 133, Piscataway, NJ 08855-1331.

The papers in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and, in the interests of timely dissemination, are published as presented and without change. Their inclusion in this publication does not necessarily constitute endorsement by the editors, the IEEE Computer Society, or the Institute of Electrical and Electronics Engineers, Inc.

IEEE Computer Society Order Number P3867
BMS Part Number CFP09090-PRT
ISBN 978-0-7695-3867-9
ISSN Number 1095-1350

Additional copies may be ordered from:

IEEE Computer Society
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314
Tel: + 1 800 272 6657
Fax: + 1 714 821 4641
<http://computer.org/cspress>
csbooks@computer.org

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: + 1 732 981 0060
Fax: + 1 732 981 9667
[http://shop.ieee.org/store/
customer-service@ieee.org](http://shop.ieee.org/store/customer-service@ieee.org)

IEEE Computer Society
Asia/Pacific Office
Watanabe Bldg., 1-4-2
Minami-Aoyama
Minato-ku, Tokyo 107-0062
JAPAN
Tel: + 81 3 3408 3118
Fax: + 81 3 3408 3553
tokyo.ofc@computer.org

Individual paper REPRINTS may be ordered at: <reprints@computer.org>

Editorial production by Bob Werner
Cover art production by Joe Daigle/Studio Productions
Printed in the United States of America by The Printing House



**IEEE Computer Society
Conference Publishing Services (CPS)**

<http://www.computer.org/cps>

2009 16th Working Conference on Reverse Engineering

WCRE 2009

Table of Contents

| | |
|--|------|
| Message from the General Chair | ix |
| Message from the Program Chairs | x |
| Organizing Committee | xii |
| Steering Committee | xiii |
| Program Committee | xiv |
| Additional Reviewers | xv |

Keynotes

| | |
|---|---|
| Beyond the Lone Reverse Engineer: Insourcing, Outsourcing and Crowdsourcing | 3 |
| <i>Margaret-Anne D. Storey</i> | |
| Legacy and Future of Data Reverse Engineering | 4 |
| <i>Jean-Luc Hainaut</i> | |

WCRE 1999 Most Influential Paper

| | |
|--|---|
| Ten Years Later, Experiments with Clustering as a Software Remodularization Method | 7 |
| <i>Nicolas Anquetil and Timothy C. Lethbridge</i> | |

Session I – Mining Software Repositories

| | |
|--|----|
| Who are Source Code Contributors and How do they Change? | 11 |
| <i>Massimiliano Di Penta and Daniel M. German</i> | |
| A Study of the Time Dependence of Code Changes | 21 |
| <i>Omar Alam, Bram Adams, and Ahmed E. Hassan</i> | |
| Relating Identifier Naming Flaws and Code Quality: An Empirical Study | 31 |
| <i>Simon Butler, Michel Wermelinger, Yijun Yu, and Helen Sharp</i> | |
| Techniques for Identifying the Country Origin of Mailing List Participants | 36 |
| <i>Ran Tang, Ahmed E. Hassan, and Ying Zou</i> | |

Session II – Dynamic Analysis

| | |
|--|----|
| NTrace: Function Boundary Tracing for Windows on IA-32 | 43 |
| <i>Johannes Passing, Alexander Schmidt, Martin von Löwis, and Andreas Polze</i> | |
| Recovering Views of Inter-System Interaction Behaviors | 53 |
| <i>Christopher Ackermann, Mikael Lindvall, and Rance Cleaveland</i> | |
| Mining Quantified Temporal Rules: Formalism, Algorithms, and Evaluation | 62 |
| <i>David Lo, Ganesan Ramalingam, Venkatesh Prasad Ranganath, and Kapil Vaswani</i> | |

Session III – Empirical Software Engineering

| | |
|---|----|
| An Exploratory Study of the Impact of Code Smells on Software Change-proneness | 75 |
| <i>Foutse Khomh, Massimiliano Di Penta, and Yann-Gaël Guéhéneuc</i> | |
| An Empirical Study on Inconsistent Changes to Code Clones at Release Level | 85 |
| <i>Nicolas Bettenburg, Weyi Shang, Walid Ibrahim, Bram Adams, Ying Zou, and Ahmed E. Hassan</i> | |
| Lexicon Bad Smells in Software | 95 |
| <i>Surafel Lemma Abebe, Sonia Haiduc, Paolo Tonella, and Andrian Marcus</i> | |

Session IV – Remodularization and Reengineering

| | |
|---|-----|
| Automatic Package Coupling and Cycle Minimization | 103 |
| <i>Hani Abdeen, Stéphane Ducasse, Houari Sahraoui, and Ilham Alloui</i> | |
| Identifying Cycle Causes with Enriched Dependency Structural Matrix | 113 |
| <i>Jannik Laval, Simon Denier, Stéphane Ducasse, and Alexandre Bergel</i> | |
| The Logical Modularity of Programs | 123 |
| <i>Daniel Ratiu, Radu Marinescu, and Jan Jürjens</i> | |
| On the Use of ADM to Contextualize Data on Legacy Source Code for Software Modernization | 128 |
| <i>Ricardo Pérez-Castillo, Ignacio García-Rodríguez de Guzmán, Orlando Ávila-García, and Mario Piattini</i> | |

Session V - Change and Defect Proneness

| | |
|---|-----|
| On the Relationship Between Change Coupling and Software Defects | 135 |
| <i>Marco D'Ambros, Michele Lanza, and Romain Robbes</i> | |
| Tracking Design Smells: Lessons from a Study of God Classes | 145 |
| <i>Stéphane Vaucher, Foutse Khomh, Naouel Moha, and Yann-Gaël Guéhéneuc</i> | |
| Bug-Inducing Language Constructs | 155 |
| <i>Javed Ferzund, Syed Nadeem Ahsan, and Franz Wotawa</i> | |
| Design Patterns and Change Proneness: A Replication Using Proprietary C# Software | 160 |
| <i>Matt Gatrell, Steve Counsell, and Tracy Hall</i> | |

Session VI – Static Analysis and Security

| | |
|---|-----|
| Automatic Static Unpacking of Malware Binaries | 167 |
| <i>Kevin Coogan, Saumya Debray, Tasneem Kaochar, and Gregg Townsend</i> | |
| Computing the Structural Difference between State-Based Models | 177 |
| <i>Kirill Bogdanov and Neil Walkinshaw</i> | |
| Extraction of Inter-procedural Simple Role Privilege Models from PHP Code | 187 |
| <i>Dominic Letarte and Ettore Merlo</i> | |

Session VII – Traceability

| | |
|---|-----|
| Traceability Recovery Using Numerical Analysis | 195 |
| <i>Giovanni Capobianco, Andrea De Lucia, Rocco Oliveto, Annibale Panichella, and Sebastiano Panichella</i> | |
| Benchmarking Lightweight Techniques to Link E-Mails and Source Code | 205 |
| <i>Alberto Bacchelli, Marco D'Ambros, Michele Lanza, and Romain Robbes</i> | |
| Domain Feature Model Recovery from Multiple Applications Using Data Access Semantics and Formal Concept Analysis | 215 |
| <i>Yiming Yang, Xin Peng, and Wenyun Zhao</i> | |

Session VIII - Program Comprehension

| | |
|---|-----|
| Characterizing Evolutionary Clusters | 227 |
| <i>Adam Vanya, Steven Klusener, Nico van Rooijen, and Hans van Vliet</i> | |
| Autumn Leaves: Curing the Window Plague in IDEs | 237 |
| <i>David Roethlisberger, Oscar Nierstrasz, and Stéphane Ducasse</i> | |
| Constructing a Resource Usage View of a Large and Complex Software-Intensive System | 247 |
| <i>Trosky Boris Callo Arias, Pierre America, and Paris Avgeriou</i> | |

Session IX – Static Analysis

| | |
|---|-----|
| Static Detection of Disassembly Errors | 259 |
| <i>Nithya Krishnamoorthy, Saumya Debray, and Keith Fligg</i> | |
| Reverse Engineering Sequence Diagrams for Enterprise JavaBeans with Business Method Interceptors | 269 |
| <i>Alexander Serebrenik, Serguei Roubtsov, Ella Roubtsova, and Mark van den Brand</i> | |
| Computing Structural Types of Clone Syntactic Blocks | 274 |
| <i>Ettore Merlo and Thierry Lavoie</i> | |
| Reverse Engineering Existing Web Service Applications | 279 |
| <i>Houda El Bouhissi and Mimoun Malki</i> | |

PhD Forum

| | |
|--|-----|
| Supporting Feature-Level Software Maintenance | 287 |
| <i>Meghan Revelle</i> | |
| Enabling the Evolution of J2EE Applications through Reverse Engineering and Quality Assurance | 291 |
| <i>Fabrizio Perin</i> | |
| Approximate Graph Matching in Software Engineering | 295 |
| <i>Sègla Kpodjedo</i> | |
| Evolving Software Systems Towards Adaptability | 299 |
| <i>Mehdi Amoui</i> | |
| SQUAD: Software Quality Understanding through the Analysis of Design | 303 |
| <i>Foutse Khomh</i> | |

Tool Demonstrations

| | |
|--|-----|
| PRECISO: A Reverse Engineering Tool to Discover Web Services from Relational Databases | 309 |
| <i>Ricardo Pérez-Castillo, Ignacio García-Rodríguez de Guzmán, Ismael Caballero, Macario Polo, and Mario Piattini</i> | |
| Recovering Class Models Stereotyped with Crosscutting Concerns | 311 |
| <i>Heitor Augustus Xavier Costa, Paulo Afonso Parreira Júnior, Valter Vieira de Camargo, and Rosângela Aparecida Delloso Penteadó</i> | |
| SHINOBI: A Tool for Automatic Code Clone Detection in the IDE | 313 |
| <i>Shinji Kawaguchi, Takano Yu Yamashina, Hidetake Uwano, Kyohei Fushida, Yasutaka Kamei, Masataka Nagura, and Hajimu Iida</i> | |
| Enhancing Quality of Code Clone Detection with Program Dependency Graph | 315 |
| <i>Yoshiki Higo and Shinji Kusumoto</i> | |
| JavaCompExt: Extracting Architectural Elements from Java Source Code | 317 |
| <i>Nicolas Anquetil, Jean-Claude Royer, Pascal André, Gilles Ardourel, Petr Hnětynka, Tomáš Poch, Dragoş Petraşcu, and Vladieiela Petraşcu</i> | |
| ConAn: A Tool for the Identification of Crosscutting Concerns in Object Oriented Systems Based on Type Hierarchy Analysis | 319 |
| <i>Mario Luca Bernardi and Giuseppe Antonio Di Lucca</i> | |

Workshops

| | |
|--|-----|
| R.E.M. 2009 - International Workshop on Reverse Engineering Models from Software Artifacts | 323 |
| <i>Leon Moonen and Tarja Systä</i> | |
| FAMOOSr 2009 - Workshop on FAMIX and Moose in Software Reengineering | 325 |
| <i>Simon Denier and Tudor Gîrba</i> | |

| | |
|---------------------------|-----|
| Author Index | 327 |
|---------------------------|-----|

PRECISO: A Reverse Engineering Tool to Discover Web Services from Relational Databases

Ricardo Pérez-Castillo, Ignacio García-Rodríguez de Guzmán, Ismael Caballero, Macario Polo and Mario Piattini

Alarcos Research Group, University of Castilla-La Mancha
Paseo de la Universidad, 4 13071, Ciudad Real, Spain
{Ricardo.PdelCastillo, Ignacio.GRodriguez, Ismael.Caballero, Mario.Piattini}@uclm.es

Abstract—There is a real need for SOA principles, such as those offering software as services, in the software industry. In this effort, databases (one of the most important artefacts in Information Systems) can be also seen as a set of services offering access to the stored information. This paper presents PRECISO, a reverse engineering tool to discover and generate Web Services automatically from relational databases. PRECISO makes it possible to modernize legacy databases by introducing them in an SOA context by means of the generated services. This tool was used in a real-life case study in the context of a software company.

Keywords: Database Reverse Engineering, Web Services, MDA, SOA and Pattern Matching.

I. INTRODUCTION

Today, organizations are increasingly forced to share more and more information as part of the basic activity in their daily operations. However, the heterogeneity of Information Systems (IS) is growing every day due to the appearance of new technological paradigms, standards, and environments, making it more and more difficult to share information [3]. Due to these facts and in order to keep their competitiveness level throughout their IS, organizations must be involved in a process of continuous renewal. Therefore, IS developers must constantly and quickly develop and maintain their products in order to meet market requirements [3]. Among all the artefacts that compose information systems, databases are possibly one of the most important elements since they contain all the organizational information and form the basis of decision-making.

This paper presents PRECISO [1], a tool for database re-engineering following the MDA (*Model-Driven Architecture*) principles [5] to extract Web Services (WS) that show the database as a set of services, offering easy access to the information. PRECISO offers several benefits: (i) it minimizes heterogeneity problems since databases can be integrated in SOA environments; (ii) it advocates the reuse of legacy databases, thus extending the lifecycle of databases; and (iii) it shortens development time because the WS generation is automatic and instantaneous.

II. THE RECOVERY PROCESS

Figure 1 depicts the database re-engineering process, which consists of three main activities broken down into several tasks. *DMR* is the first activity to create a PSM (*Platform-Specific Model*) which represents the input relational database. The *DMR-1* task recovers metadata from the database and builds the PSM according to an SQL-92 metamodel based on [2]. The *DMR-2* task simultaneously discovers the potential services by means of pattern matching [4]. Figure 2 shows the patterns recognized in the database schema and the associated service templates.

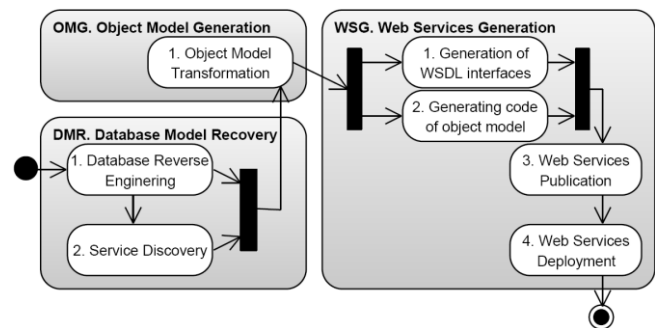


Figure 1. The Web Services recovery process

The second activity is *OMG*, which generates an object model from the previous database schema model. Thus, the *OMG-1* task transforms the PSM into a PIM (*Platform-Independent Model*) according to the UML2 metamodel [6]. Finally, the third activity is *WSG*, which generates the WS to manage the input database. The *WSG-1* task builds the service interfaces in WSDL (*Web Services Description Language*) from services discovered in *DMR-2*. This set of interfaces is the target PSM since it represents the WS platform. At the same time, *WSG-2* generates the source code related to the object model. This code supports the business layer of WS. Then *WSG-3* writes out the source code of the WS. After that, *WSG-4* carries out the Web Services deployment. Finally, the WS are ready to be used.

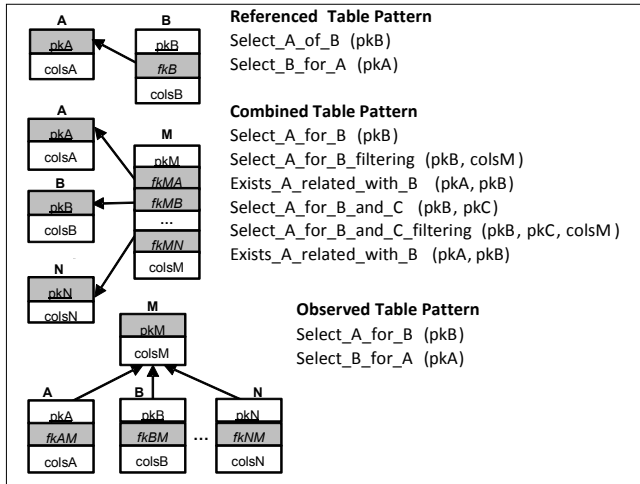


Figure 2. Service patterns in database schema

III. THE ARCHITECTURE

The PRECISO tool automates several tasks in the proposed recovery process (see Figure 1); but it also addresses other necessary issues such as remote database connection, connections to databases from different vendors, project management, graphic display of involved models, testing, reporting, and so on. The proposed architecture, taking into account the previous challenges, is shown in Figure 3.

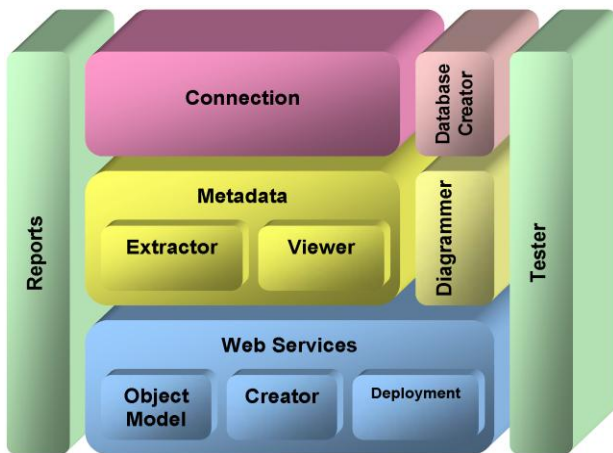


Figure 3. Architecture of PRECISO

The most important features of PRECISO are: (i) *'project-oriented nature'* because it manages a repository with all recovery information; (ii) *'partitioned and ordered process generation'* since PRECISO generates WS by means of individual, but complementary modules; (iii) *'usage of standards'* such as SQL-92, XMI (*XML Metadata Interchange*) and WSDL. (Therefore, PRECISO can be integrated with other reverse engineering tools in the market); and (iv) *'high development process coverage'* since it supports metadata extraction, model generation,

graphic visualization of models, model transformations, editing and publication of WS, deployment, testing, reporting, and so on.

IV. CONCLUSIONS

This paper has presented PRECISO, a tool that carries out a re-engineering process following the MDA standard to discover and generate WS from relational databases. PRECISO presents three key advantages: (i) it integrates relational databases into SOA environments; (ii) it extends the lifecycle of legacy databases; and (iii) it accelerates Web developments. In addition, PRECISO has been validated through a case study since it was used in an industry project to develop a corporate Web portal from a legacy database. The tool selectively published and deployed of the generated WS. The percentage of services that was published to support the functionalities of the Web portal was 30%. This percentage represents 73 services out of 245 candidate services in total. The WS supports all the information needs of the portal in an SOA context, and thus the staff could put all of their effort into the development of the Web interface.

The future extensions of this work will make more in-depth analysis to establish more patterns in database schemes in order to define more service templates.

ACKNOWLEDGMENTS

This work has been supported by the *FPU Spanish Program*; by the R+D projects funded by *JCCM: ALTAMIRA (PII2109-0106-2463)*, *INGENIO (PAC08-0154-9262)* and *PRALIN (PAC08-0121-1374)*; and *MITOS (TC20091098)* funded by the *University of Castilla-La Mancha*.

REFERENCES

- [1] Alarcos, R.G. PRECISO. A Reverse Engineering Tool to Discover Web Services from Relational Databases. 2008 [cited 7/30/2009]; Available from: <http://alarcos.esi.uclm.es/per/rpdelcastillo/PRECISO/PRECISO.html>.
- [2] Calero, C., "An Ontological Approach To Describe the SQL:2003 Object-Relational Features". *Journal of Computer Standards and Interfaces*, 2006. 28(6): p. 695-713.
- [3] Di Lucca, G.A., A.R. Fasolino, and P. Tramontana, "Reverse engineering Web applications: the WARE approach". *Journal of Software Maintenance and Evolution: Research and Practice*, 2004. 16(1-2): p. 71 - 101.
- [4] García, I., M. Polo, and M. Piattini, "Using Model-Driven Pattern Matching to derive functionalities in Models", in *9th International Conference on Software Engineering and Knowledge Engineering*. 2007: Boston, USA. p. 529-534.
- [5] Miller, J. and J. Mukerji, "MDA Guide" Version 1.0.1. 2003: OMG.
- [6] OMG. UML (Unified Modeling Language). Superstructure specification. Version 2.2. 2009 02-02-2009 [cited 24-06-2009]; Available from: <http://www.omg.org/spec/UML/2.2/Superstructure/PDF/>.