

Lecture Notes in Computer Science

The LNCS series reports state-of-the-art results in computer science research, development, and education, at a high level and in both printed and electronic form. Enjoying tight cooperation with the R&D community, with numerous individuals, as well as with prestigious organizations and societies, LNCS has grown into the most comprehensive computer science research forum available.

The scope of LNCS, including its subseries LNAI and LNBI, spans the whole range of computer science and information technology including interdisciplinary topics in a variety of application fields. The type of material published traditionally includes

- proceedings (published in time for the respective conference)
- post-proceedings (consisting of thoroughly revised final full papers)
- research monographs (which may be based on outstanding PhD work, research projects, technical reports, etc.)

More recently, several color-cover sublines have been added featuring, beyond a collection of papers, various added-value components; these sublines include

- tutorials (textbook-like monographs or collections of lectures given at advanced courses)
- state-of-the-art surveys (offering complete and mediated coverage of a topic)
- hot topics (introducing emergent topics to the broader community)

In parallel to the printed book, each new volume is published electronically in LNCS Online.

Detailed information on LNCS can be found at www.springer.com/lncs

Proposals for publication should be sent to LNCS Editorial, Tiergartenstr. 17, 69121 Heidelberg, Germany
E-mail: lncs@springer.com

ISSN 0302-9743

ISBN 978-3-540-76291-1




9 783540 762911

Lecture Notes in
Computer Science

LNCS

LNAI

LNBI

 springer.com

Hainaut et al. (Eds.)



LNCS
4802

Advances in Conceptual Modeling –
Foundations and Applications

ER 2007
Workshops

LNCS 4802

Advances in Conceptual Modeling – Foundations and Applications

ER 2007 Workshops CMLSA, FP-UML,
ONISW, QoIS, RIGiM, SeCoGIS
Auckland, New Zealand, November 2007, Proceedings

 Springer

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

University of California, Los Angeles, CA, 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, Saarbrücken, Germany

Jean-Luc Hainaut Elke A. Rundensteiner
Markus Kirchberg Michela Bertolotto
Mathias Brochhausen Yi-Ping Phoebe Chen
Samira Si-Said Cherfi Martin Doerr Hyoil Han
Sven Hartmann Jeffrey Parsons Geert Poels
Colette Rolland Juan Trujillo Eric Yu
Esteban Zimányi (Eds.)

Advances in Conceptual Modeling – Foundations and Applications

ER 2007 Workshops CMLSA, FP-UML,
ONISW, QoIS, RIGiM, SeCoGIS
Auckland, New Zealand, November 5-9, 2007
Proceedings

 Springer

Volume Editors

Jean-Luc Hainaut E-mail: jean-luc.hainaut@fundp.ac.be
Elke A. Rundensteiner E-mail: rundenst@cs.wpi.edu
Markus Kirchberg E-mail: M.Kirchberg@massey.ac.nz
Michela Bertolotto E-mail: michela.bertolotto@ucd.ie
Mathias Brochhausen E-mail: mathias.brochhausen@fomis.uni-saarland.de
Yi-Ping Phoebe Chen E-mail: phoebe@deakin.edu.au
Samira St-Said Cerf E-mail: sisaid@cnam.fr
Martin Doerr E-mail: martin@ics.forth.gr
Hyoil Han E-mail: hhan@ischool.drexel.edu
Sven Hartmann E-mail: s.hartmann@massey.ac.nz
Jeffrey Parsons E-mail: jeffreyp@mun.ca
Geert Poels E-mail: geert.poels@ugent.be
Colette Rolland E-mail: rolland@univ-paris1.fr
Juan Trujillo E-mail: jtrujillo@dlsi.ua.es
Eric Yu E-mail: eric.yu@utoronto.ca
Esteban Zimányi E-mail: ezimanyi@ulb.ac.be

Library of Congress Control Number: 2007938044

CR Subject Classification (1998): H.2, H.4, H.3, F.4.1, D.2, C.2.4, I.2, J.1

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

ISSN 0302-9743

ISBN-10 3-540-76291-4 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-76291-1 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

springer.com

© Springer-Verlag Berlin Heidelberg 2007

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12181322 06/3180 5 4 3 2 1 0

Preface

The 26th International Conference on Conceptual Modelling in Auckland, New Zealand, hosted six workshops which allowed participants to focus their presentations and discussions on advanced topics that cannot easily fit the general conference scope.

Thirteen good quality proposals were received and nine were selected. Due to the similarity of their scope, two pairs were suggested to merge, leading to seven proposals. One workshop attracted fewer submissions than expected, so that its selected papers were integrated into the conference. Finally, six workshops were kept. Interestingly, four of them (FP-UML, ONISW, QoIS, SeCoGIS) were a sequel of workshops that were held in the last few years, while two were new (CMLSA, RIGIM), exhibiting both the maturity and the innovation of the workshops.

Following the call for papers, we received 114 complete submissions, from which 40 quality papers were selected, giving an acceptance rate of 35% (a fairly standard score for workshops).

The following six workshops were organized:

- **Conceptual Modelling for Life Sciences Applications (CMLSA 2007)**, chaired by Yi-Ping Phoebe Chen and Sven Hartmann. This workshop addressed the specific challenges posed by the large data volumes, the complexity and the data and software heterogeneity involved by life science applications.
- **Foundations and Practices of UML (FP-UML 2007)**, chaired by Juan Trujillo and Jeffrey Parsons. The third edition of this workshop gathered researchers and practitioners on topics related to data warehouses, security, model transformation, state diagrams development and model quality.
- **Ontologies and Information Systems for the Semantic Web (ONISW 2007)**, chaired by Mathias Brochhausen, Martin Doerr and Hyoil Han. The second edition of this workshop focused on the potential and actual roles of ontologies in Web information systems. In particular, the papers addressed ontology extraction from texts, ontology interoperability, automatic semantic annotation and domain modeling.
- **Quality of Information Systems (QoIS 2007)**, chaired by Samira St-Said Cerf and Geert Poels. The papers of the third edition of this workshop dealt with quality models, model understandability and (at last!) quality assessment of scientific conferences.
- **Requirements, Intentions and Goals in Conceptual Modeling (RIGIM 2007)**, chaired by Colette Rolland and Eric Yu. The workshop goal was to explore the relations between requirements engineering and conceptual modeling. The selected papers particularly addressed such topics as model-driven approaches, the use of *i**, refinement of goal-based methodologies, elicitation techniques and visualization.

– Semantic and Conceptual Issues in Geographic Information Systems (SeCo-GIS 2007), chaired by Esteban Zimányi and Michela Bertolotto. The fourth edition of this workshop included presentations on data interoperability, spatial data warehouses, spatial data querying, data quality, conceptual modeling, spatio-temporal aspects of moving objects as well as special spatial applications.

We are particularly grateful to all the workshop organizers, the Program Committee members and the authors for the time and effort spent to guarantee the high quality of the programs. Special thanks are due to Markus Kirchberg who took charge of most of the editorial work of these proceedings.

November 2007

Jean-Luc Hainaut
Elke A. Rundensteiner

ER 2007 Conference Organization

General Chair

Bernhard Thalheim (Christian Albrechts University Kiel, Germany)

Program Committee Co-chairs

Christine Parent (University of Lausanne, Switzerland)
Klaus-Dieter Schewe (Massey University, New Zealand)
Veda C. Storey (Georgia State University, USA)

Organization Chair

Gillian Dobbie (University of Auckland, New Zealand)

Steering Committee Liaison

Tok Wang Ling (National University of Singapore, Singapore)

Publicity chair

Markus Kirchberg (Massey University, New Zealand)

Workshop Co-chairs

Jean-Luc Hainaut (University of Namur, Belgium)
Elke A. Rundensteiner (Worcester Polytechnic Institute, USA)

Tutorial Co-chairs

Sven Hartmann (Massey University, New Zealand)
Alberto H.F. Laender (UFMG, Brazil)

Panel Chair

John F. Roddick (Flinders University, Australia)

Industrial Chair

John Grundy (University of Auckland, New Zealand)

Demonstration and Poster Chair

Leszek Maciaszek (Macquarie University, Australia)

Treasurers

Patricia Rood (University of Auckland, New Zealand)
Stephen W. Liddle (Brigham Young University, USA)

Organized By

Massey University, New Zealand
The University of Auckland, New Zealand

Sponsored By

The ER Institute

In Cooperation With

ACM SIGMIS
ACM SIGMOD

ERR 2007 Workshop Organization

ERR 2007 Workshop Co-chairs

Jean-Luc Hainaut, University of Namur, Belgium
Elke A. Rundensteiner Worcester Polytechnic Institute, USA

CMISA 2007 – International Workshop on Conceptual Modelling for Life Sciences Applications

Workshop Chairs

Yi-Ping Phoebe Chen, Deakin University, Australia
Sven Hartmann, Massey University, New Zealand

Publicity Chair

Markus Kirchberg, Massey University, New Zealand

Program Committee

Isabelle Bichindaritz, USA	Jingchu Luo, China
Leonard N. Blosberg, New Zealand	Victor Maojo, Spain
Jake Chen, USA	Sudha Ram, USA
Alexei Drummond, New Zealand	Keun Ho Ryu, Korea
Silke Eckstein, Germany	Amandeep S. Sidhu, Australia
Amar Nath Gupta, USA	Tanya Soboleva, New Zealand
David Hansen, Australia	Theodoros Topaloglou, Canada
Peter Hunter, New Zealand	Haixun Wang, USA
Nikola Kasabov, New Zealand	Jing Wang, New Zealand
Dirk Labudde, Germany	Xiaofang Zhou, Australia
Sebastian Link, New Zealand	Esteban Zimányi, Belgium
Huiqing Liu, USA	

External Referees

Claudia Täubner, Germany
Henning Köhler, New Zealand

FP-UML 2007 – International Workshop on Foundations and Practices of UML

FP-UML 2007 was organized within the framework of the following projects: META-SIGN (TIN2004-00779) from the Spanish Ministry of Education and Science and DADS (PBC-05-012-2) from the Castilla-La Mancha Ministry of Science and Technology (Spain).

Workshop Co-chairs

Juan Trujillo, University of Alicante, Spain
Jeffrey Parsons, Memorial University of Newfoundland, Canada

Program Committee

Doo-Hwan Bae, South Korea
Michael Blaha, USA
Cristina Cachero, Spain
Tharam Dillon, Australia
Gillian Dobbie, New Zealand
Brian Dobing, Canada
Dirk Draheim, Germany
Joerg Evermann, New Zealand
Eduardo Fernández, Spain
Jens Lechtenböcker, Germany
Tok Wang Ling, Singapore
Pericles Loucopoulos, UK
Hui Ma, New Zealand

Andreas L. Opdahl, Norway
Oscar Pastor, Spain
Witold Pedrycz, Canada
Mario Piattini, Spain
Ivan Porres, Finland
Colette Rolland, France
Matti Rossi, Finland
Manuel Serrano, Spain
Keng Siau, USA
Il-Yeol Song, USA
Ambrosio Toval, Spain
Antonio Vallecillo, Spain
Panos Vassiliadis, Greece

External Referees

M. Kirchberg
A. Tretiakov

O. Thonggoom
G. Abraham

ONISW 2007 – International Workshop on Ontologies and Information Systems for the Semantic Web

Workshop Co-chairs

Mathias Brochhausen, Saarland University, Germany
Martin Doerr, Foundation for Research and Technology - Hellas, Greece
Hyoil Han, Drexel University, USA

Program Committee

Yuan An, Canada
Sofia J. Athenikos, USA
Boualem Benattallah, Australia
Mathias Brochhausen, Germany
Namyoon Choi, USA
Cristian Cocos, Germany
Martin Doerr, Greece
Ramez Elmasri, USA
Fabien Gandon, France
Raul Garcia-Castro, Spain

Hyoil Han, USA
Mary Elizabeth Jones, USA
Aneesh Krishna, Australia
Werner Kuhn, Germany
Sang-Koo Lee, Korea
SeungJin Lim, USA
Dimitris Plexousakis, Greece
Barry Smith, USA and Germany
Krishnaprasad Thirunarayan, USA

QoIS 2007 – International Workshop on Quality of Information Systems

Workshop Co-chairs

Samira Si-Said Cherfi, CEDRIC-CNAM, France
Geert Poels, Ghent University, Belgium

Steering Committee

Jacky Akoka, CEDRIC-CNAM & INT, France
Mokrane Bouzeghoub, PRISM, University of Versailles, France
Isabelle Comyn-Wattiau, Conservatoire National des Arts et Métiers and ESSEC, France
Marcela Genero, Universidad de Castilla-La Mancha, Spain
Jeffrey Parsons, Memorial University of Newfoundland, Canada
Geert Poels, Ghent University, Belgium
Keng Siau, University of Nebraska, USA
Bernhard Thalheim, University of Kiel, Germany

Program Committee

Jacky Akoka, France
Laure Berti-Equille, France
Andrew Burton-Jones, Canada
Tiziana Catarci, Italy
Corinne Cauvet, France
Isabelle Comyn-Wattiau, France
Marcela Genero, Spain
Paul Johannesson, Sweden
Jacques Le Maître, France
Jim Nelson, USA
Jeffrey Parsons, Canada
Oscar Pastor, Spain

Houari Sabraoui, Canada
Farida Semmak, France
Keng Siau, USA
Guttorm Sindre, Norway
Samira Si-Said Cherfi, France
Monique Snoeck, Belgium
Il-Yeol Song, USA
David Tegarden, USA
Bernhard Thalheim, Germany
Dimitri Theodoratos, USA
Juan Trujillo, Spain

External Referees

Ki Jung Lee, USA

RIGIM 2007 – International Workshop on Requirements, Intentions and Goals in Conceptual Modeling**Workshop Co-Chairs**Colette Rolland, Université Paris1 Panthéon Sorbonne, France
Eric Yu, University of Toronto, Canada**Program Committee**

Daniel Amyot, Canada
Mikio Aoyoma, Japan
Ian Alexander, UK
Aybuke Arum, Australia
Franck Barbier, France
Daniel Berry, Canada
Sjaak Brinkkemper, The Netherlands
Lawrence Chung, USA
Luiz Cysneiros, Canada
Eric Dubois, Luxembourg
Vincenzo Gervasi, Italy
Aditya K. Ghose, Australia
Peter Haumer, USA
Zhi Jin, China
Aneesh Krishna, Australia
John Krogstie, Norway
Lin Liu, China

Peri Loucopoulos, UK
John Mylopoulos, Canada
Selmin Nurcan, France
Baslar Nuseibeh, UK
Andreas Opdahl, Norway
Barbara Pernici, Italy
Klaus Pohl, Germany
Jolita Ralyte, Switzerland
Bjorn Regnell, Sweden
Carnille Salinesi, France
Motoshi Saeki, Japan
Pnina Soffer, Israel
Carine Souveyet, France
Leon Sterling, Australia
Yair Wand, Canada
Roel Wieringa, The Netherlands

External Referees

Arosha Bandara, UK
Chiara Francalanci, Italy
Joy Garfield, UK

Kim Lauenroth, Germany
Thorsten Weyer, Germany

SeCoGIS 2007 – International Workshop on Semantic and Conceptual Issues in Geographic Information Systems**Workshop Co-chairs**

Esteban Zimányi, Université Libre de Bruxelles, Belgium
Michela Bertolotto, University College Dublin, Ireland

Program Committee

Gennady Andrienko, Germany
Natalia Andrienko, Germany
Yvan Bédard, Canada
David Bennett, USA
Alex Borgida, USA
Patrice Boursier, France
Bénédicte Bucher, France
James Carswell, Ireland
Christophe Claramunt, France
Eliseo Clementini, Italy
Maria Luisa Damiani, Italy
Clodoveu Davis, Brazil
Max Egenhofer, USA
Fernando Ferri, Italy
Andrew Frank, Austria
Anders Friis-Christensen, Italy
Antony Galton, UK
Bo Huang, Canada
Marinos Kavouras, Greece

Werner Kuhn, Germany
Sergei Levashkin, Mexico
Ki-Joune Li, South Korea
Thérèse Libourel, France
Jose Macedo, Switzerland
Peter van Oosterom, The Netherlands
Dimitris Papadias, China
Fabio Porto, Switzerland
Dieter Pfoser, Greece
Ricardo Rodrigues Ciferri, Brazil
Andrea Rodriguez, Chile
Sylvie Servigne-Martin, France
Stefano Spaccapietra, Switzerland
Emmanuel Stefanakis, Greece
Kerry Taylor, Australia
Christelle Vangenot, Switzerland
Antonio Miguel Vieira Monteiro, Brazil
Agnès Voisard, Germany

Table of Contents

CMLSA 2007 – International Workshop on Conceptual Modelling for Life Sciences Applications	
Preface to CMLSA 2007	1
<i>Yi-Ping Phoebe Chen and Sven Hartmann</i>	
Knowledge Discovery in Life Sciences	
Pattern Recognition of Single-Molecule Force Spectroscopy Data	3
<i>Dirk Labudde, Annalisa Marsico, K. Tanuj Sapra, and Michael Schroeder</i>	
Massive Protein Structural Property Explorations Using New Indexing Mechanism	14
<i>Yu-Feng Huang, Chia-Chen Chang, and Chien-Kang Huang</i>	
Data Integration and Exchange in Health Informatics	
Data Access and Management in ACCT: Tools to Solve Syntactic and Semantic Heterogeneities Between Clinical and Image Databases	24
<i>Luis Martín, Erwin Bonsma, Alberto Anguita, Jeroen Vrijnsen, Miguel García-Remesal, José Crespo, Manolis Tsiknakis, and Victor Maojo</i>	
Ontology-Based Data Integration in Data Logistics Workflows	34
<i>Olivier Cuvé and Stefan Jablonski</i>	
Model-Driven Development Based Transformation of Stereotyped Class Diagrams to XML Schemas in a Healthcare Context	44
<i>Eladio Domínguez, Jorge Lloret, Beatriz Pérez, Áurea Rodríguez, Ángel L. Rubio, and María A. Zapata</i>	
Conceptual Modelling for Biological Systems	
An Extendable System for Conceptual Modeling and Simulation of Signal Transduction Pathways	54
<i>Silke Eckstein and Claudia Täubner</i>	
Toward an Ontological Database for Subcellular Neuroanatomy	64
<i>Amarnath Gupta, Stephen D. Larson, Christopher Condit, Sandeep Gupta, Lisa Fong, Li Chen, and Maryann E. Martone</i>	

Seed-Based Generation of Personalized Bio-ontologies for Information Extraction	74
<i>Cui Tao and David W. Embley</i>	
FP-UML 2007 – International Workshop on Foundations and Practices of UML	
Preface to FP-UML 2007	85
<i>Juan Trujillo and Jeffrey Parsons</i>	
Improving the Use of UML Diagrams	
Developing State Diagrams Using a State Specialization Technique	86
<i>Il-Yeol Song and Ki Jung Lee</i>	
Quality Dependencies Among Use Case Models and Sequence Diagrams Developed by Novice Systems Analysts	96
<i>Narasimha Bolloju and Vijayan Sugumar</i>	
M-BPSeq: A Method for Security Requirement Elicitation from a UML 2.0 Business Process Specification	106
<i>Alfonso Rodríguez, Eduardo Fernández-Medina, and Mario Piattini</i>	
Model Transformations and Extensions	
Applying Model Transformation By-Example on Business Process Modeling Languages	116
<i>Michael Strommer, Marion Marzok, and Manuel Wimmer</i>	
Extending OCL to Ensure Model Transformations	126
<i>François Lagarde, François Terrier, Charles André, and Sébastien Gérard</i>	
A UML Profile for Modeling Data Warehouse Usage	137
<i>Veronika Stefanov and Beate List</i>	
ONISW 2007 – International Workshop on Ontologies and Information Systems for the Semantic Web	
Preface to ONISW 2007	149
<i>Mathias Brochhausen, Martin Doerr, and Hyoil Han</i>	
A Method for Semi-automatic Creation of Ontologies Based on Texts	150
<i>Luiz C.C. Carvalho and Edson Satoshi Gomi</i>	
Enriching OWL with Instance Recognition Semantics for Automated Semantic Annotation	160
<i>Yihong Ding, David W. Embley, and Stephen W. Liddle</i>	

Making Web Users' Domain Models Explicit by Applying Ontologies ...	170
<i>Tarnio Robal, Hele-Mai Haav, and Ahto Kahja</i>	
Provability-Based Semantic Interoperability Via Translation Graphs	180
<i>Joshua Taylor, Andrew Shilliday, and Selmer Bringsjord</i>	
QoIS 2007 – International Workshop on Quality of Information Systems	
Preface to QoIS 2007	191
<i>Samira Si-Said Cherfi and Geert Peels</i>	
PQM vs. BPQM: Studying the Tailoring of a General Quality Model to a Specific Domain	192
<i>Corral Calero, Cristina Cachero, Julio Córdoba, and M^{ra} Angeles Moraga</i>	
An Ontological Approach for the Quality Assessment of Computer Science Conferences	202
<i>Maria Aparecida M. Souto, Mariusa Warpechowski, and José Palazzo M. de Oliveira</i>	
Using Practitioners for Assessing the Understandability of UML Statechart Diagrams with Composite States	213
<i>José A. Cruz-Lemus, Marcela Genero, Sandro Morasca, and Mario Piattini</i>	
RIGIM 2007 – International Workshop on Requirements, Intentions and Goals in Conceptual Modelling	
Preface to RIGIM 2007	223
<i>Colette Rolland and Eric Yu</i>	
Keynote	
An Ontology for Requirements	224
<i>John Mylopoulos, Iuan J. Jureta, and Stéphane Faulner</i>	
Requirements and Goals – Methods	
GOORE: Goal-Oriented and Ontology Driven Requirements Elicitation Method	225
<i>Masayuki Shibata, Haruhiko Kaiga, and Motoshi Sasaki</i>	
Early Prioritisation of Goals	235
<i>Sarah Hutton</i>	

Goal-Aligned Requirements Generation	245
<i>Kousik Senkar Ramasubramaniam and Raman Venkatachar</i>	
A Model-Driven Goal-Oriented Requirement Engineering Approach for Data Warehouses	255
<i>Jose-Norberto Mazón, Jesús Pardillo, and Juan Trujillo</i>	
Visually Effective Goal Models Using KAOS	265
<i>Raimundas Matulevičius and Patrick Heymans</i>	
Agent Based Executable Conceptual Models Using i* and CASO	276
<i>Aniruddha Dasgupta, Aneesh Krishna, and Aditya K. Ghose</i>	
Requirements and Goals – Concepts	
Achieving, Satisficing, and Excelling	286
<i>Ivan J. Jureta, Stéphane Faulkner, and Pierre-Yves Schobbens</i>	
On the Adequacy of i* Models for Representing and Analyzing Software Architectures	296
<i>Gemma Grau and Xavier Franch</i>	
Extending Argumentation to Goal-Oriented Requirements Engineering	306
<i>Ibrahim Habli, Weihang Wu, Katrina Atwood, and Tim Kelly</i>	
SeCoGIS 2007 – International Workshop on Semantic and Conceptual Issues in Geographic Information Systems	
Preface to SeCoGIS 2007	317
<i>Esteban Zimányi and Michela Bertolotto</i>	
Moving Objects	
Modeling Historical and Future Spatio-temporal Relationships of Moving Objects in Databases	318
<i>Reasey Praing and Markus Schneider</i>	
Towards a Semantic Spatial Model for Pedestrian Indoor Navigation ...	328
<i>Edgar-Philipp Stoffel, Bernhard Lorenz, and Hans Jürgen Ohlbach</i>	
Modeling Collaborative Semantics with a Geographic Recommender ...	338
<i>Christoph Schlieder</i>	
Dynamically Traveling Web Service Clustering Based on Spatial and Temporal Aspects	348
<i>Yang Hao, Chen Junliang, Meng Xiangyu, and Qiu Bingyu</i>	
Advances in Conceptual Modelling for GIS	
A Graph-Oriented Model and Query Language for Events	358
<i>Miguel Romero and M. Andrea Rodríguez</i>	
PLR Partitions: A Conceptual Model of Maps	368
<i>Mark McKenney and Markus Schneider</i>	
A Conceptual Framework to Support Semantic Interoperability of Geospatial Databases	378
<i>Tarek Sboui, Yvan Bédard, Jean Brodeur, and Thierry Badard</i>	
Integrity Constraints and Approximate Reasoning	
On Languages for the Specification of Integrity Constraints in Spatial Conceptual Models	388
<i>Mehrdad Salehi, Yvan Bédard, Mir Abolfazl Mostafavi, and Jean Brodeur</i>	
Approximate Queries by Relaxing Structural Constraints in GIS	398
<i>Arianna D'Ulizia, Fernando Ferri, and Patrizia Grifoni</i>	
Ensuring the Semantic Correctness of Complex Regions	409
<i>Mark McKenney, Alejandro Pauty, Reasey Praing, and Markus Schneider</i>	
Author Index	419

Using Practitioners for Assessing the Understandability of UML Statechart Diagrams with Composite States

José A. Cruz-Lemus¹, Marcela Genero¹, Sandro Morasca², and Mario Piattini¹

¹ ALARCOS Research Group
Department of Information Technologies and Systems
Indra-UCLM Research and Development Institute
University of Castilla-La Mancha

Paseo de la Universidad, 4 – 13071 Ciudad Real, Spain

{JoseAntonio.Cruz, Marcela.Genero, Mario.Piattini}@uclm.es

² Dipartimento di Scienze Della Cultura, Politiche e dell'Informazione

Università degli Studi dell'Insubria, Como, Italy
Sandro.Morasca@uninsubria.it

Abstract. We have carried out a family of empirical studies to investigate whether the use of composite states improves the understandability of UML statechart diagrams. Our hypothesis derived from conventional wisdom, which says that hierarchical modeling mechanisms are helpful to master a system's complexity. We carried out three studies that have gradually evolved in the size of the UML statechart models, the type of subjects (students vs. professionals), the familiarity of the subjects with the domains of the diagrams, and other factors. In this work we briefly review the first and second studies and present the third one, performed with practitioners as experimental subjects. Surprisingly, our results do not seem to show that the use of composite states improves the understandability of UML statechart diagrams.

1 Introduction

Models help us understand a complex problem and its potential solutions through abstraction. This is why software systems can benefit from using models and modeling techniques [21]. The Model-Driven Development (MDD) paradigm [1] focuses the effort of development on the design of models, rather than coding. Correspondingly, the focus of software quality assurance is shifting from system implementation towards system modeling. So, to be useful and effective, an engineering model must possess the following five key quality characteristics to a sufficient degree [21]: abstraction, understandability, accuracy, predictiveness and inexpensiveness.

In this paper, we focus on understandability because it is recognized as one of the main factors influencing maintainability, and it is well-recognized that a large part of the effort invested in the development of any software product is devoted to maintenance [19]. More specifically, we focus on the understandability of UML statechart diagrams, since UML has now become the de facto standard for modeling software systems and UML statechart diagrams have become an important technique for describing the dynamic aspects of a software system [7].

UML statechart diagrams are also considered as one of the most important UML diagrams [3] and should be used by practitioners as a starting point to train UML newcomers. Although the use of this kind of diagram is not critical for many systems, they are very useful for their intended purpose, have low redundancy and provide new helpful information in some systems [8], specially in real-time systems [9].

We have previously studied [5] the effect that some UML constructs may have on the understandability of UML statechart diagrams, including composite states, which are an important construct of the UML statechart diagrams metamodel [17] and a fundamental modeling abstraction mechanism. UML statechart diagrams with composite states extend finite state machines to facilitate the description of highly complex behaviors [13]. A composite state is a state that contains other states within it. Composite states are believed to be useful to model the behavior of complex classes, as we can group the simple states that are part of a larger common one, and so they help structure a system and master a system's complexity by dividing the system in smaller, less complex parts. However, it is necessary to provide empirical support to this belief to show if it is actually true and, if so, under what conditions.

Thus, based on the common use of hierarchical structures in modeling techniques for making easier to reason about a system, both for those who write the design of a system and those who read it [11, 20, 23], we hypothesized that grouping simple states that are highly related into a composite state could help improve the understandability of a diagram. We have performed a family of experiments to test this hypothesis, and in this paper we present the third empirical study corresponding to this family. We have gradually modified and improved the design and execution of these experiments, so as to alleviate some threats to the validity of the different studies that are part of it.

In Section 2 of this paper, we provide a roadmap of the family of experiments that we have performed. Section 3 explains the main features and conclusions of the two first members of the family of experiments that has been performed. Section 4 comments in detail the experimental process performed in order to carry out the third experiment of the family. Finally, section 5 summarizes the main lessons learned from this family of experiments and the future work planned to perform.

2 The Family of Experiments

An experiment can be viewed as part of common families of studies, rather than being an isolated event [2]. Common families of studies can contribute to devising important and relevant hypotheses that may not be suggested by individual experiments.

Fig. 1 shows the chronology of the family of experiments we have carried out in our study on the understandability of UML statechart diagrams.

In the initial studies (E1, R1, E2, and R2), we used students as experimental subjects. The tasks to be performed did not require high levels of industrial experience, so we thought that this experiment could be considered appropriate [2, 12]. In addition, working with students implies a set of advantages, such as the fact that the prior knowledge of the students is rather homogeneous, the availability of a large number of subjects [26], and the possibility of testing experimental design and initial hypotheses [22].

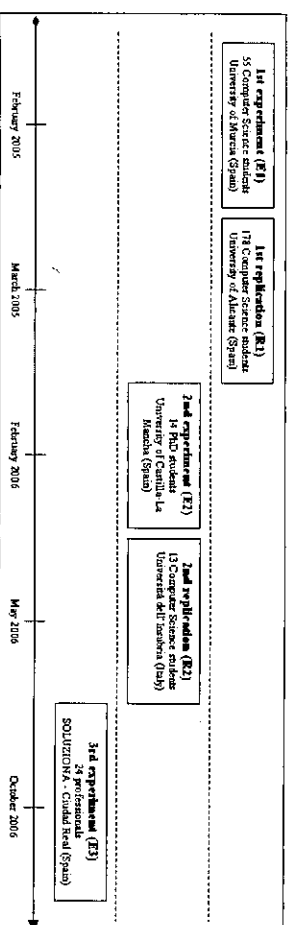


Fig. 1. Chronology of the family of experiments

The main difference between these studies and the third experiment (E3) presented in this paper, lies both in having professionals as experimental subjects, and also having renewed and improved materials and tasks.

In studies E1 and R1, we used the understandability effectiveness variable, defined as the ability to understand the presented material correctly. In studies E2, R2 and E3, we added two new variables related to the CTML, retention and transfer. Retention is defined as the comprehension of material being presented. Transfer is the ability to use knowledge gained from the material to solve related problems not directly answerable from it.

To measure all these variables, we have used three separate tests based on questionnaires. Each understandability characteristic measure is computed as the proportion of correct answers provided in each test. In this work, we will call these values *UEffec* for understandability effectiveness, *UTrans* for transfer, and *UReten* for retention, respectively.

We carried out the experimental processes based on well-known guidelines for empirical studies [14, 15, 28] and we used SPSS [24] to perform all the statistical analyses. In all of our studies, we set a statistical significance threshold $\alpha = 0.05$.

3 First and Second Empirical Studies

In this section, we briefly outline the main features (see Table 1 and Table 2) and conclusions reached after the performance of the two first members of the family of experiments. The complete details about them can be found in [4] and [6].

Table 1. E1 And R1 main features

Location	Universities of Murcia, Spain (E1) and Alicante, Spain (R1)
Date	February (E1) and March (R1) 2005
Subjects	55 (E1) and 178 (R1) Computer Science students
Dependent Variable	Understandability of UML statechart diagrams, measured through the measure of <i>UEffec</i> .
Independent Variables	The use or not of composite states (CS) in the diagrams.
Alternative Hypotheses	H_{1a} : using composite states improves the <i>UEffec</i> in subjects when trying to understand an UML statechart diagram.

The materials and tasks to be performed in the first experiment and its replication (E1 and R1) were quite simple and the background knowledge of the subjects was not advanced.

The results obtained in these experiments indicated that using composite states did not improve the understandability of UML statechart diagrams. This assertion was completely against our original hypothesis, so we considered the study as preliminary and decided to perform further empirical studies for accepting or definitively reject our hypothesis.

Table 2. E2 and R2 main features

Location	Universities of Castilla-La Mancha, Spain (E2) and Insubria, Italy (R2)
Date	February (E2) and May (R2) 2006
Subjects	14 PhD students (E2) and 13 Computer Science students (R2)
Dependent Variable	Understandability of UML statechart diagrams, measured through the measures of <i>UEffect</i> , <i>UTrans</i> , and <i>UReten</i> .
Independent Variables	The use or not of composite states (CS) in the diagrams.
Alternative-Hypotheses	H1a: using composite states improves the <i>UEffect</i> in subjects when trying to understand an UML statechart diagram. H1b: using composite states improves the <i>UTrans</i> in subjects when trying to understand an UML statechart diagram. H1c: using composite states improves the <i>UReten</i> in subjects when trying to understand an UML statechart diagram.

With respect to the second experiment and its replication (E2 and R2), the materials and tasks were improved, especially with the use of the Cognitive Theory for Multimedia Learning (CTML) [16] for assessing the complete set of variables of the experimental design. More about this theory can be found in [6, 10, 16].

The results obtained with the *UEffect* variable agreed with the ones obtained in the E1 and R1, i.e., those UML statechart diagrams modeled without using composite states were easier to understand. Besides, the values obtained for the new variables introduced in this study, retention and transfer, were in almost all cases middle values, neither especially high nor low. This way, no strong assertions could be made about them.

So, we decided revising and improving the materials and tasks to perform in a new experiment, trying to throw light to these uncertain conclusions.

4 Third Experiment

In this section, we explain the process we followed when we carried out the third member of the family of experiments (E3).

The experiment was carried out in the facilities of Indra Company, located in Ciudad Real, Spain. Indra currently holds a top position in the market of professional

services in software, with a sales volume which is close to 800 million euros and, after a long period of expansion, the company has spread to over 28 countries in 4 different continents. The company has recently reached maturity level 3, according to the CMMi model, and the achievement of level 4 is planned for the current year 2007.

4.1 Design

In this study, we based on the design of a digital watch [27], through a model which has a size and complexity representative of a real-life case. The working hypotheses and part of the procedure to follow were similar to E2 and R2. Table 3 outlines the main features of E3.

Table 3. E3 main features

Location	Ciudad Real (Spain)
Date	October 2006
Subjects	24 professionals (12 in each group)
Dependent Variable	Understandability of UML statechart diagrams, measured through the measures of <i>UEffect</i> , <i>UTrans</i> , and <i>UReten</i> .
Independent Variables	The use or not of composite states (CS) in the diagrams.
Alternative-Hypotheses	H1a: using composite states improves the <i>UEffect</i> in subjects when trying to understand an UML statechart diagram. H1b: using composite states improves the <i>UTrans</i> in subjects when trying to understand an UML statechart diagram. H1c: using composite states improves the <i>UReten</i> in subjects when trying to understand an UML statechart diagram.

4.2 E3 Procedure

The experiment was divided into two sessions, in the afternoon and morning of two consecutive days.

In order to have a homogenous knowledge background in the subjects, the first session began with a seminar about "Dynamic Modeling with UML." Twenty-five professionals attended the first session, and they were provided with a summary of the main concepts of dynamic aspects in modeling in general and in UML in particular. The last part of the seminar focused on the UML statechart diagrams, although there was not an explicit mention which made the subjects guess the relation between the seminar and the subsequent experiment.

After the seminar, the instructor explained several examples related with the test that the subjects were performing in that session (Test 0). These examples consisted of questions about navigation through several statechart diagrams. Then, the subjects performed Test 0. This test was used for balancing the subjects into groups depending on their knowledge and performance.

This first session lasted approximately two hours. After it, all of Test 0 questionnaires were corrected and the subjects were grouped into two groups depending on their results.

The subjects were ordered according to the number of correct answers and the time spent on the questionnaire [18]. After that, those subjects who occupied an odd rank were assigned to Group A, and the others to Group B. Thus, we obtained two balanced groups as Table 4 shows.

Table 4. Means and standard deviations for groups after Test 0

Group	N	Correct Answers		Time	
		Mean	Std. Dev.	Mean	Std. Dev.
A	13	5.3846	1.3409	628.62	122.88
B	12	5.4583	1.0967	620.67	135.18

One of the subjects who had been assigned to Group A did not show up for Session 2. We decided to not rearrange the groups, which would now have the same number of subjects.

Session 2 was composed of three phases. In phase 1, the subjects were informed that they had been grouped depending on their performance in the test collected the day before. After that, they received one UML statechart diagram and a copy of Test 1. To avoid possible learning effects, we adopted a balanced between-subjects and blocked design, i.e. each subject was assigned only one diagram. The subjects of Group A received a diagram modelled using composite states and those in Group B received exactly the same system but modelled without using composite states.

As in E2 and R2, this test was used for measuring the UEffec of the model. The questions in Test 1 covered all the different parts of the diagram so that we could make sure that all parts of the diagram had been covered by the subjects before we removed them. This phase lasted for 25 minutes.

Then, all diagrams and tests were collected and Tests 2 and 3 were handed out. Test 2 was used for measuring the UReten variable and consisted of a fill-in-the-blanks text with 10 gaps that the subjects had to complete in order to build the text with the specifications of the system. The subjects had 15 minutes for this phase.

Test 3 was used for measuring the UTrans variable and consisted of a list of 6 tasks to perform based on the information taken from the diagram. The subjects had 35 minutes to solve them.

All the questions and tasks to perform in the different tests were exactly the same for both groups.

After the end of Test 3, all the materials were collected and the subjects were handed out a final questionnaire, to collect the subjects' impressions about the difficulty of the tests and the main positive and negative points that they had found during the experiment.

Fig. 2 graphically describes all this process.

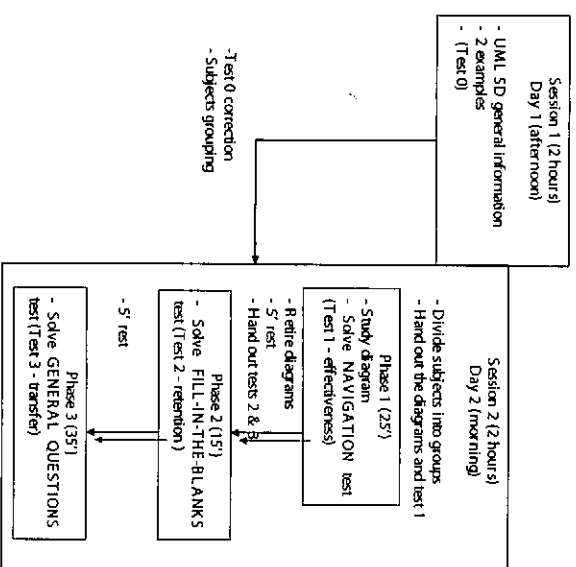


Fig. 2. E3 procedure schema

All the experimental material and data of the complete family of experiments can be found at the website <http://alarcos.inf-cr.uclm.es/CSExperiments/>

4.3 E3 Data Analysis and Interpretation

As in the previous studies, we carried out an analysis of the descriptive statistics of the data. Table 5 presents the descriptive statistics for the measures of the dependent variables studied in E3.

We can observe how, in this case, the results obtained for the understandability effectiveness and transfer variables are higher when the subjects worked with the diagram modelled with composite states, while the retention variable was higher in the diagram modelled without using composite states.

Table 5. Descriptive statistics for E3

CS	With (n=12)				Without (n=12)			
	Stat.	Mean	Std. Dev.	Min. Max.	Mean	Std. Dev.	Min. Max.	
UEffec	0.750000	0.1446	0.5	0.9	0.641667	0.1564	0.4	0.9
UTrans	0.369048	0.2079	0.071	0.786	0.261906	0.1677	0.036	0.571
UReten	0.775000	0.1390	0.5	1.0	0.891667	0.0515	0.8	1.0

After removing the outlier values, those which were too high and low and had fallen out of the range of typical values, we performed a t-test for testing the three hypotheses shown in Table 3 (see Table 6).

Table 6. E3 t-test results

Domain	UEffec		UTrans		UReten	
	t	sig.	t	sig.	t	sig.
Digital Watch	1.7617	0.092	1.3895	0.179	-3.6093	0.002

The results indicate that not using composite states positively affects the retention of the subjects while using composite states (weakly) improves UEffec, since the result is actually not statistically significant with our 0.05 threshold, but would be significant with a 0.1 threshold.

Regardless of the use of composite states, the subjects obtained a high level on the retention test and a low value in the transfer one. The CTML suggests that in this case, there has been only fragmented understandability but not memorization. In our opinion, the increased difficulty of the diagrams and the tasks might have produced this effect.

5 Lessons Learned and Future Work

In this work, we have presented a controlled experiment, last member of a family of empirical studies, for assessing how the use of composite states in an UML statechart diagram affects the understandability of the diagram.

We have followed the steps suggested in the Empirical Software Engineering, beginning our study with students for testing the original designs [25] and gradually improving the materials used and the experience of the subjects until performing the last study with a group of professionals.

After reviewing the obtained results in the family of experiments, we reach the following conclusions:

1. Our first idea, and the one most commonly accepted in the Software Engineering field, was that using composite states helps make a UML statechart be more understandable. The results of our empirical study do not support this idea. In most of the cases, the understandability effectiveness of the experimental subjects was better when working with systems modelled without using composite states. Nevertheless, in our last experiment, in which the subjects were professionals and hence more experienced, and the model was quite complex, the results for the understandability effectiveness were somewhat better for the subjects that used a model with composite states.
2. There is not a clear effect either in using or in not using composite states on the concepts of transfer and retention introduced by the CTML [16]. The increasing difficulty of the diagrams used and the task proposed to the subjects might have affected them, by making the diagrams more difficult to be understood.

Unexpectedly, despite performing a family of empirical studies, the evidence we collected does not support our initial hypothesis as clearly supported as we thought, and composite states do not seem to be highly useful for UML statechart diagrams understandability.

We have focused on understandability and we are planning to study some other quality aspects, such as modifiability, that could be affected by the presence of composite states. Tasks related to modifiability will probably be more complex and we will try to check whether the use of composite states makes them easier to perform.

Acknowledgements

This research is part of the MECENAS project (PBI06-0024) financed by "Consejería de Ciencia y Tecnología de la Junta de Comunidades de Castilla-La Mancha" and the ESPINGE project (TIN2006-15175-C05-05) and the CALIPSO Network (TIN20005-24055-E) supported by the "Ministerio de Educación y Ciencia" (Spain). The research presented in this paper has been partially funded by the IST project "QualiPSo," sponsored by the EU in the 6th FP (IST-034763), the FIRB project "ARTDECO," sponsored by the Italian Ministry of Education and University, and the project "La qualità nello sviluppo software," sponsored by the Università degli Studi dell'Insubria.

The authors would like to thank the staff in SOLUZIONA - Ciudad Real for their time and understanding during the preparation and performance of this experiment.

References

1. Atkinson, C., Kühne, T.: Model Driven Development: a Metamodeling Foundation. IEEE Transactions on Software Engineering 20, 36-41 (2003)
2. Basili, V., Shull, F., Lanubile, F.: Building Knowledge through Families of Experiments. IEEE Transactions on Software Engineering 25, 456-473 (1999)
3. Bolloju, N., Leung, F.S.K.: Assisting Novice Analysts in Developing Quality Conceptual Models with UML. Communications of the ACM 49(7), 108-112 (2006)
4. Cruz-Lemus, J.A., Genero, M., Manso, M.E., Piattini, M.: Evaluating the Effect of Composite States on the Understandability of UML Statechart Diagrams. In: Briand, L.C., Williams, C. (eds.) MoDELS 2005. LNCS, vol. 3713, pp. 113-125. Springer, Heidelberg (2005)
5. Cruz-Lemus, J.A., Genero, M., Piattini, M.: Metrics for UML Statechart Diagrams. In: Genero, M., Piattini, M., Calero, C. (eds.) Metrics for Software Conceptual Models, Imperial College Press, UK (2005)
6. Cruz-Lemus, J.A., Genero, M., Piattini, M., Morasca, S.: Improving the Experimentation for Evaluating the Effect of Composite States on the Understandability of UML Statechart Diagrams. In: Proceedings of 5th ACM-IEEE International Symposium on Empirical Software Engineering, Rio de Janeiro, Brazil, pp. 9-11 (2006)
7. Denger, C., Ciolkowski, M.: High Quality Statecharts through Tailored, Perspective-Based Inspections. In: Proceedings of 29th EUROMICRO Conference New Waves in System Architecture, Belek, Turkey, pp. 316-325 (2003)
8. Dobing, B., Parsons, J.: How UML is Used. Communications of the ACM 49(5), 109-113 (2006)
9. Erickson, J., Siau, K.: Can UML Be Simplified? Practitioner Use of UML in Separate Domains. In: EMMSAD 2007. Proceedings of Twelfth International Workshop on Exploring Modeling Methods in Systems Analysis and Design, Trondheim, Norway, pp. 89-98 (2007)

10. Germino, A., Wand, Y.: Evaluating Modeling Techniques based on Models of Learning. *Communications of the ACM* 46(10), 79–84 (2003)
11. Ghezzi, C., Jazayeri, M., Mandrioli, D.: *Fundamentals of Software Engineering*. Prentice Hall, Englewood Cliffs (1992)
12. Höst, M., Regnell, B., Wohlin, C.: Using Students as Subjects - a Comparative Study of Students & Professionals in Lead-Time Impact Assessment. In: *EASE 2000. Proceedings of 4th Conference on Empirical Assessment & Evaluation in Software Engineering*, Keele, UK, pp. 201–214 (2000)
13. Hu, Z., Shanz, S.M.: Explicit Modeling of Semantics Associated with Composite States in UML Statecharts. *Automated Software Engineering* 13(4), 423–467 (2006)
14. Juristo, N., Moreno, A.: *Basics of Software Engineering Experimentation*. Kluwer Academic Publishers, Boston (2001)
15. Kitchenham, B., Pfleeger, S., Pickard, L., Jones, P., Hoaglin, D., El-Emani, K., Rosenberg, J.: Preliminary Guidelines for Empirical Research in Software Engineering. *IEEE Transactions on Software Engineering* 28(8), 721–734 (2002)
16. Mayer, R.E.: *Multimedia Learning*. Cambridge University Press, Cambridge (2001)
17. OMG, UML 2.0 - 2nd Revised Submission, Object Management Group (2003)
18. Otero, M.C., Dolado, J.J.: Evaluation of the Comprehension of the Dynamic Modeling in UML. *Information and Software Technology* 46(1), 35–53 (2004)
19. Pigoski, T.: *Practical Software Maintenance*. Wiley Computer Publishing, New York, USA (1997)
20. Pressman, R.: *Ingeniería del Software*. McGraw-Hill, New York (1993)
21. Selic, B.: The Pragmatics of Model-Driven Development. *IEEE Software* 20(5), 19–25 (2003)
22. Sjøberg, D.I.K., Hamay, J.E., Hansen, O., Kampenes, V., Karahasanovic, A., Liborg, N.K., Rekdal, A.C.: A Survey of Controlled Experiments in Software Engineering. *IEEE Transactions on Software Engineering* 31(9), 733–753 (2005)
23. Sommerville, I.: *Software Engineering*. Addison Wesley, Reading (1995)
24. SPSS, SPSS 12.0, Syntax Reference Guide, SPSS Inc.: Chicago, USA (2003)
25. Tichy, W.F.: Hints for Reviewing Empirical Work in Software Engineering. *Empirical Software Engineering* 5, 309–312 (2000)
26. Vereist, J.: The Influence of the Level of Abstraction on the Evolvability of Conceptual Models of Information Systems. In: *ISESE 2004. Proceedings of 3rd International Symposium on Empirical Software Engineering*, Redondo Beach, USA, pp. 17–26 (2004)
27. Webb, K.: *Xtalon Digital Watch Project* (2006)
28. Wohlin, C., Runeson, P., Hast, M., Ohlsson, M.C., Regnell, B., Westlen, A.: *Experimentation in Software Engineering: an Introduction*. Kluwer Academic Publishers, Boston (2000)