Editors

Vladimir Klebanov
Karlsruhe Institute of Technology
Institute for Theoretical Informatics
Am Fasanengarten 5, 76131 Karlsruhe, Germany
Email: klebanov@kit.edu

Bernhard Beckert
Karlsruhe Institute of Technology
Institute for Theoretical Informatics
Am Fasanengarten 5, 76131 Karlsruhe, Germany
Email: beckert@kit.edu

Armin Biere
Institute for Formal Models and Verification
Johannes Kepler University
Altenbergerstr. 69, 4040 Linz, Austria
Email: biere@jku.at

Geoff Sutcliffe
Department of Computer Science
University of Miami
P.O. Box 248154, Coral Gables, FL 33124-4245, USA
Email: geoff@cs.miami.edu

Copyright © 2012 for the individual papers by the papers’ authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.
Preface

This volume contains the proceedings of the 1st International Workshop on Comparative Empirical Evaluation of Reasoning Systems (COMPARE 2012), held on June 30th, 2012 in Manchester, UK, in conjunction with the International Joint Conference on Automated Reasoning (IJCAR).

It has become accepted wisdom that regular comparative evaluation of reasoning systems helps to focus research, identify relevant problems, bolster development, and advance the field in general. Benchmark libraries and competitions are two popular approaches to do so. The number of competitions has been rapidly increasing lately. At the moment, we are aware of about a dozen benchmark collections and two dozen competitions for reasoning systems of different kinds. It is time to compare notes.

What are the proper empirical approaches and criteria for effective comparative evaluation of reasoning systems? What are the appropriate hardware and software environments? How to assess usability of reasoning systems, and in particular of systems that are used interactively? How to design, acquire, structure, publish, and use benchmarks and problem collections?

The aim of the workshop was to advance comparative empirical evaluation by bringing together current and future competition organizers and participants, maintainers of benchmark collections, as well as practitioners and the general scientific public interested in the topic.

We wish to sincerely thank all the authors who submitted their work for consideration. All submitted papers were peer-reviewed, and we would like to thank the Program Committee members as well as the additional referees for their great effort and professional work in the review and selection process. Their names are listed on the following pages. We are deeply grateful to our invited speakers—Leonardo de Moura (Microsoft Research) and Cesare Tinelli (University of Iowa)—for accepting the invitation to address the workshop participants. We thank Sarah Grebing for her help in organizing the workshop and compiling this volume.

June 2012

Vladimir Klebanov
Bernhard Beckert
Armin Biere
Geoff Sutcliffe
Program Committee

Bernhard Beckert  Karlsruhe Institute of Technology, Germany
Christoph Benzmüller  Free University Berlin, Germany
Dirk Beyer  University of Passau, Germany
Armin Biere  Johannes Kepler University Linz, Austria
Vinay Chaudhri  SRI International, USA
Koen Claessen  Chalmers Technical University, Sweden
Alberto Griggio  Fondazione Bruno Kessler, Italy
Marieke Huisman  University of Twente, the Netherlands
Radu Iosif  Verimag/CNRS/University of Grenoble, France
Vladimir Klebanov  Karlsruhe Institute of Technology, Germany
Rosemary Monahan  National University of Ireland Maynooth
Michal Moskal  Microsoft Research, USA
Jens Otten  University of Potsdam, Germany
Franck Pommereau  University of Évry, France
Sylvie Putot  CEA-LIST, France
Olivier Roussel  CNRS, France
Albert Rubio  Universitat Politècnica de Catalunya, Spain
Aaron Stump  University of Iowa, USA
Geoff Sutcliffe  University of Miami, USA

Program Co-Chairs

Bernhard Beckert  Karlsruhe Institute of Technology, Germany
Armin Biere  Johannes Kepler University Linz, Austria
Vladimir Klebanov  Karlsruhe Institute of Technology, Germany
Geoff Sutcliffe  University of Miami, USA

Organising Committee

Bernhard Beckert  Karlsruhe Institute of Technology, Germany
Sarah Grebing  Karlsruhe Institute of Technology, Germany
Vladimir Klebanov  Karlsruhe Institute of Technology, Germany

Additional Referees

Sarah Grebing
# Table of Contents

## Abstracts of Invited Talks

Regression Tests and the Inventor’s Dilemma ............................... 1  
  Leonardo de Moura

Introducing StarExec: a Cross-Community Infrastructure for Logic  
Solving ......................................................... 2  
  Aaron Stump, Geoff Sutcliffe, and Cesare Tinelli

## Contributed Papers

Evaluating the Usability of Interactive Verification Systems ............. 3  
  Bernhard Beckert and Sarah Grebing

Broadening the Scope of SMT-COMP: the Application Track ............. 18  
  Roberto Bruttomesso and Alberto Griggio

A Simple Complexity Measurement for Software Verification and  
Software Testing .................................................. 28  
  Zheng Cheng, Rosemary Monahan, and James Power

Benchmarking Static Analyzers ............................................. 32  
  Pascal Cuoq, Florent Kirchner, and Boris Yakobowski

The 2nd Verified Software Competition: Experience Report ............. 36  
  Jean-Christophe Filliâtre, Andrei Paskevich, and Aaron Stump

On the Organisation of Program Verification Competitions ............... 50  
  Marieke Huisman, Vladimir Klebanov, and Rosemary Monahan

Challenges in Comparing Software Verification Tools for C ............. 60  
  Florian Merz, Carsten Sinz, and Stephan Falke

Behind the Scene of Solvers Competitions: the “evaluation” Experience 66  
  Olivier Roussel

## Author Index

................................................................. 78