

A List-Machine Benchmark for Mechanized Metatheory

Andrew W. Appel · Robert Dockins · Xavier Leroy

Received: 28 July 2009 / Accepted: 8 April 2011 / Published online: 3 May 2011
© Springer Science+Business Media B.V. 2011

Abstract We propose a benchmark to compare theorem-proving systems on their ability to express proofs of compiler correctness. In contrast to the first POPLmark, we emphasize the connection of proofs to compiler implementations, and we point out that much can be done without binders or alpha-conversion. We propose specific criteria for evaluating the utility of mechanized metatheory systems; we have constructed solutions in both Coq and Twelf metatheory, and we draw conclusions about those two systems in particular.

Keywords Theorem proving · Proof assistants · Program proof · Compiler verification · Typed machine language · Metatheory · Coq · Twelf

1 How to Evaluate Mechanized Metatheories

The POPLmark challenge [4] aims to compare the usability of several automated proof assistants for mechanizing the kind of programming-language proofs that

Andrew W. Appel and Robert Dockins were supported in part by NSF Grant CNS-0910448.

A. W. Appel · R. Dockins
Princeton University, Princeton, NJ 08540, USA

A. W. Appel
e-mail: appel@princeton.edu

R. Dockins
e-mail: rdockins@princeton.edu

X. Leroy (✉)
INRIA Paris-Rocquencourt, B.P. 105, 78153 Le Chesnay, France
e-mail: xavier.leroy@inria.fr