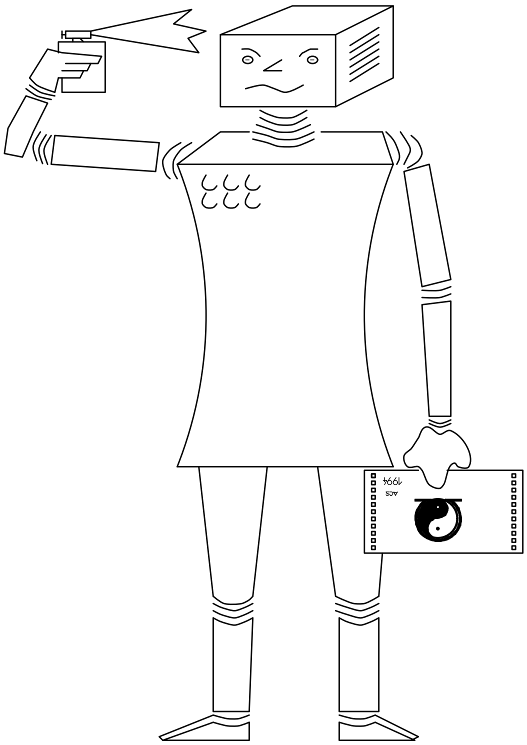


Guaranteeing Correctness through the Communication of Checkable Proofs

(or: Would You Really Trust an Automated Reasoning System?)

Xiaorong Huang Manfred Kerber
Michael Kohlhase Dan Nesmith
Jörn Richts

Published as: In David Basin, Fausto Giunchiglia, Matt Kaufmann, eds., *Proceedings of the CADE-Workshop on Correctness and Metatheoretic Extensibility of Automated Reasoning Systems*, Nancy, France, p.31–33, 1994.

Guaranteeing Correctness through the Communication of Checkable Proofs*

(or: Would You Really Trust an Automated Reasoning System?)

Xiaorong Huang Manfred Kerber Michael Kohlhase
Dan Nesmith Jörn Richts
Fachbereich Informatik, Universität des Saarlandes
Postfach 1150, 66041 Saarbrücken, Germany
{huang|kerber|kohlhase|nesmith|richts}@cs.uni-sb.de

> Die Mathematik ist *nicht* nach ihrem Gegenstand (etwa: Raum und Zeit, Formen der inneren Anschauung, Lehre vom Zählen und Messen u. dergl.) zu charakterisieren, sondern, wenn man ihren ganzen Umfang erschöpfen will, allein durch ihr eigentümliches Verfahren, den Beweis. *Ernst Zermelo*

1 The Notion of Proof and Correctness

Mathematics generally enjoys the prestige of being the *correct* scientific discipline *par excellence*. This reputation comes from the requirement that every claim must be justified by a rigorous proof. Looked at a bit more closely, however, the general notion of proof is by no means simple and fixed but is used to cover diverse concepts and plays different roles. In traditional mathematical practice, proofs are not given in terms of single calculus rules but at a level of abstraction that conveys the main ideas. This procedure is based on the conviction that a detailed logic-level proof could be generated if necessary, which, however, would be too boring. In addition, most mathematicians have no interest in "formal" (logical-level) proofs. The correctness of such relatively informal proofs is usually guaranteed by a social process of critical reviewing. It turns out nevertheless that this social process often fails to reach its goal: in most cases this is only caused by minor and reparable errors, but from time to time a false theorem is assumed to have been proven. The history of Euler's polyhedron theorem is a well-known story of such repeated falsification and patching [4].

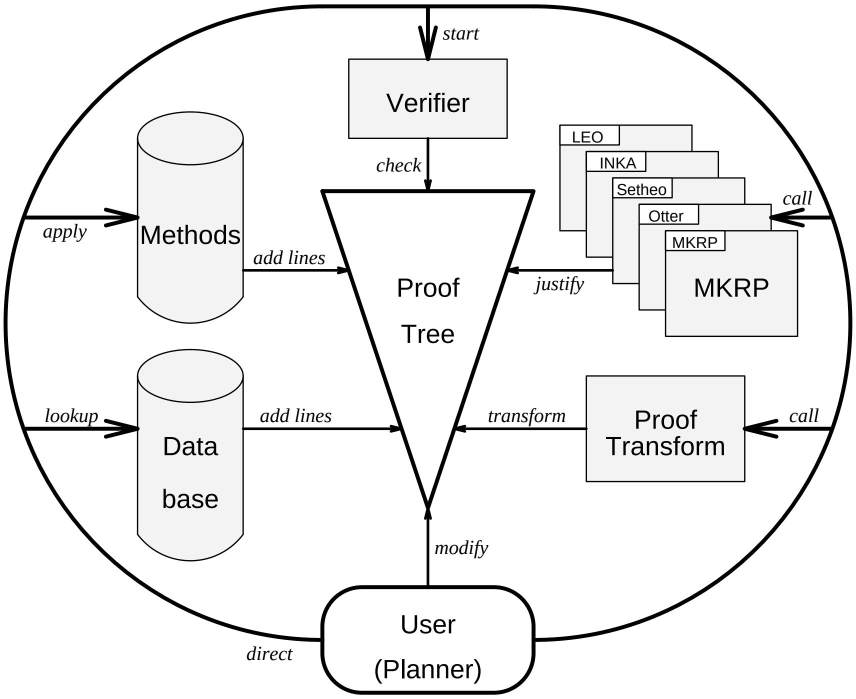
The development of mathematical logic and in particular of automated reasoning systems is an attempt to achieve a new quality of correctness [1]. The philosophy behind this enterprise lies in the belief that the machine guarantees the correctness

*This position paper reflects an ongoing discussion in our group.

by carrying out only correct deduction steps on a calculus level. As pointed out in [5] two different positions are taken towards the notion proof, namely the *realist* and the *nominalist* position. A realist accepts abstract properties of proofs, in particular he is satisfied with the evidence of the existence of a proof in order to accept the truth of a theorem. A proof for a nominalist, on the other hand, makes only sense with respect to a particular calculus, hence he only accepts concrete proofs formulated in this calculus. The advantage of adopting the realist position is that reasoning systems can be built (and meta-theoretically extended) without bothering about the concrete construction of proofs. The advantage of the nominalist position is that it preserves the tradition that proofs can be communicated. Only the nominalist position guarantees the correctness of machine generated proofs without violating an essential of the traditional notion of proof, namely their communicability. Furthermore explicit proofs can be checked by simple proof checkers and this seems to be the only way to ensure the correctness of proofs generated by large computing systems, which are inevitably error-prone.

2 Correctness in Deduction Systems

In accordance with the two philosophical positions there are two ways to use automated reasoning systems: as trustworthy black box or as a system that produces communicable proofs. Based on the discussion in the previous section we follow the latter approach in Ω-MKRP [3]. In order to achieve both machine checked correctness and the readability of proofs we have chosen the following architecture:



Different external components incorporated within our proof development environment can manipulate the partial proof tree representing the current natural deduction proof. Since we cannot rely on the correctness of all of them, the final proof is checked by a simple verifier equipped with a fixed set of natural deduction rules. Our approach requires that each component protocols its results in the format of natural deduction. The final natural deduction proof together with plan-level information provides an adequate starting point for presenting the proof in a way that is familiar to mathematicians [2]. We view this transformation and checking procedure as complementary to the meta-theoretical guarantee of correctness and not as a substitute. The former guarantees the ultimate correctness of proofs generated with the help of complex systems, while the latter raises the reliability of such systems. In practice, our approach is only feasible if the incorporated components are highly trustworthy.

In the automated reasoning community powerful deduction systems have been built and will continue to be built. In order to ensure the correctness of proofs generated by these systems in a social process it is an urgent need to set up a standard format of proofs such that the proofs can be easily checked by different independently developed proof checkers.

References

[1] Nicolaas Govert de Bruijn. A survey of the project AUTOMATH. In J.P. Seldin and J.R. Hindley, editors, *To H.B. Curry - Essays on Combinatory Logic, Lambda Calculus and Formalism*, pages 579–606. Academic Press, London, United Kingdom, 1980.

[2] Xiaorong Huang. Reconstructing proofs at the assertion level. In Alan Bundy, editor, *Proceedings of the 12th CADE*, Nancy, 1994. Springer Verlag, Berlin, Germany. forthcoming.

[3] Xiaorong Huang, Manfred Kerber, Michael Kohlhase, Erica Melis, Dan Nesmith, Jörn Richts, and Jörg Siekmann. Ω-MKRP: A proof development environment. In Alan Bundy, editor, *Proceedings of the 12th CADE*, Nancy, 1994. Springer Verlag, Berlin, Germany. forthcoming.

[4] Imre Lakatos. *Proofs and Refutations*. Cambridge University Press, 1976.

[5] Francis Jeffry Pelletier. The philosophy of automated theorem proving. In John Mylopoulos and Ray Reiter, editors, *Proceedings of the 12th IJCAI*, pages 538–543, Sydney, 1991. Morgan Kaufmann, San Mateo, California, USA.