The THFTPTP Project – An Infrastructure for Typed Higher-order Form Automated Theorem Proving Marie Curie International Incoming Fellowship Grant Agreement PIIF-GA-2008-219982

Project Report – Scientific

Dr. Christoph Benzmüller (Host)
Saarland University

Prof. Dr. Geoff Sutcliffe (Researcher)
University of Miami

1 Introduction

1.1 Project Overview

There is a well established infrastructure that supports research, development, and deployment of first-order Automated Theorem Proving (ATP) systems, stemming from the Thousands of Problems for Theorem Provers (TPTP) problem library. This infrastructure includes the TPTP itself, the TPTP language and SZS result ontology, the Thousands of Solutions from Theorem Provers (TSTP) solution library, various tools associated with the libraries, and the CADE ATP System Competition (CASC). This infrastructure has been central to the impressive progress that has been made in the development of high performance first-order ATP systems. The state of the art in higher-order ATP is not as advanced as that of first-order ATP. While there are several effective interactive systems for reasoning in higher-order logic, there has been limited automation. Critically, research and development has not been supported by a commonly accepted infrastructure that provides leverage for progress leading to effective and successful application. **The completed THFTPTP project has developed an infrastructure that supports research and development of automated theorem proving in higher-order logic.** The effect of the completed research will be to support research, development, and deployment of higher-order ATP systems, so that they can be used as effective components of academic and industrial processes.

1.2 Summary of Accomplishments

The completed THFTPTP project has developed an infrastructure that supports research and development of automated theorem proving in Church's simple type theory. The software components designed and developed are:

- A higher-order TPTP language to express problems and solutions in higher-order logic.
- A collection of higher-order test problems in the TPTP.
- A result and output ontology for higher-order ATP, extending the existing SZS ontology.
- A collection of higher-order problem solutions in the TSTP.
- Tools for preparing, processing, and analyzing higher-order ATP problems and solutions.

The completed project has been the topic of, or has contributed to, paper publications, grant proposals, presentations, and events:

- Two conference papers directly about the project.
- Two workshop papers about the project and supporting topics.

- Five other papers describing work in which the new TPTP THF infrastructure was tested and employed.
- Three grant proposals that were strongly influenced by the project.
- Seven presentations related to the project.
- Six conferences at which either the host or researcher promoted use of the TPTP and the THF format.
- The new THF division of the CADE ATP System Competition (CASC)

The completed project has had significant impact:

- Significant scientific and technical impact: efective support for the development of new and more powerful ATP systems for higher-order logic, and their application to a range of existing and new application domains.
- Seeding of multiple new research directions and projects: projects that will develop new principles and practice of automated theorem proving for higher-order logic, the use of higher-order logic for theorem proving in non-classical logics, and further development of the TPTP THF infrastructure.
- Quantifiable transfer of knowledge and expertise to Europe: to the host, researchers and students at the host university, and to eleven further institutions in Europe. Further unquantified development of expertise and knowledge follows from the interaction of the host and researcher with colleagues in Europe and beyond.

2 Deliveries

2.1 Software

The THFTPTP project has delivered all the software products described in the project proposal, and additionally the project has contributed to the development of five fully automated ATP systems for higher-order logic in the THF format. The software delivered is as follows:

- A collection of THF problems in the TPTP. TPTP v4.0.0 includes 2729 THF problems in 14 problem domains. The TPTP library (currently v4.0.0) is available from `http://www.tptp.org`. The TPTP can also be browsed online from `http://www.tptp.org`. The major contributions of THF problems were:
  - A translation of Jutting's AUTOMATH formalization [18] of the well known Landau book [12].
  - An export of the TPS problem library [].
  - Security problems in access control logic and authorization logic, initially encoded in modal logic, and subsequently encoded in higher-order logic [5, 11].
  - Problems about Ramsey numbers, some of which are open in the mathematics community.
  - Problems concerning higher-order abstract syntax, encoded in higher-order logic [10].
  - Problems in dependently typed set theory [9].
  - Modal logic problems that have been encoded in higher-order logic.
  - Translations of problems from the Intuitionistic Logic Theorem Proving (ILTP) problem library [13].
- A complete machine processable BNF specification of the THF language. This is distributed as part of the TPTP problem library, and is available online at `http://www.tptp.org/TPTP/SyntaxBNF.html`. It has been used to generate the SyntaxBNF family of parsers for THF format data. The parsers are available as stand-alone products, and are also available online in the SystemB4TPTP interface at `http://www.tptp.org/cgi-bin/SystemB4TPTP`.

- Documentation for the extended TPTP. This includes manually produced components such as the online technical manual, available at `http://www.tptp.org/TPTP/TR/TPTPTR.shtml`, and mechanically generated parts such as those in the TPTP distribution's `Documents` directory. These documents are also available online at `http://www.tptp.org/cgi-bin/SeeTPTP?Category=Documents`.
- The higher-order extension of the TPTP2X tool, including core parsing and pretty printing of THF data, transformations from intuitionistic and modal logics to THF, and export of TPTP format THF problems in TPS, Isabelle, Twelf, OmDoc, and S-expression format. TPTP2X is part of the TPTP distribution, and is available online in the SystemB4TPTP interface.
- The higher-order extension of the JJParser library that provides support for reading, manipulating, and outputing TPTP format data. The JJParser library is part of the TPTP ServiceTools package, available from `http://www.tptp.org/ServiceTools.tgz`.
- The higher-order extension of the Java-based parser for reading, manipulating, and outputting TPTP format data. The Java-based parser is part of the TPTP ServiceTools package.
- The higher-order extension of the TPTP4X tool, to parse and pretty print THF data. This is available online in the SystemB4TPTP interface.
- The higher-order extension of the IDV proof presentation tool. IDV is in the ServiceTools package, and also online in the SystemOnTSTP interface at `http://www.tptp.org/cgi-bin/SystemOnTSTP`.
- The higher-order extension of the GDV proof verification tool. GDV is in the ServiceTools package, and also available online in the SystemOnTSTP interface.
- A type-checking tool for THF problems. The tool is implemented by export of THF problems in the Twelf syntax, and submission to the Twelf tool. This was joint work with Dr. Florian Rabe from the KWARC Research Group, Jacobs University in Bremen, with technical support from Dr. Carsten Schürmann from the Programming, Logics, and Semantics Group, IT University of Copenhagen. This tool is available online in the SystemB4TPTP interface.
- The TPS 3.080227G1d ATP system for THF problems. The system was implemented by using the export of THF problems in TPS syntax, and developing of a fully automated strategy-scheduling version of the TPS ATP system. This was joint work with Dr. Chad Brown from the Programming System Group in the Institut für Informatik at Saarland University. This system is available online in the SystemOnTPTP interface at `http://www.tptp.org/cgi-bin/SystemOnTPTP`.
- The IsabelleP, IsabelleM, and IsabelleN ATP systems for THF problems. The systems were implemented by using the export of THF problems in Isabelle syntax, and developing of a fully automated strategy-scheduling version of the Isabelle ATP system. IsabelleP is a theorem prover, while IsabelleM and IsabelleN are higher-order model finders. This was joint work with Dr. Makarius Wenzel, Dr. Stefan Berghofer, and Jasmin Blanchette from the Theorem Proving Group, Institut für Informatik, Technische Universität München.
- The LEO-II ATP system for THF problems. The system directly reads THF format problems and produces THF format proofs. The development of LEO-II is research of the host of the THFTPTP project, and is thus tightly linked to the project. There has been significant mutual benefit from the linkage, in terms of testing, debugging, and deployment.

2.2 Papers

The main work of the project was conducted in the funded period from July 2008 to September 2009. However, some work was done immediately after the project was granted in winter 2007/2008. This section lists all project related papers since winter 2007/2008.

Papers on the new TPTP THF infrastructure, and supporting topics: [16, 15, 8, 7, 14]

Papers in which the new TPTP THF infrastructure was tested and employed: [3, 2, 5, 4, 6, 1]

2.3 The THF Division of the CADE ATP System Competition

The CADE ATP System Competition (CASC) is an annual evaluation of fully automatic, classical logic ATP systems [17] - it is the world championship for such systems. In CASC-22 (the fourteenth CASC, held in Montreal, Canada, in August 2009), the THF (Typed Higher-order Form) division was added as a demonstration division. This new division was based on the products to the THFTPTP project:

- The THF problems collected and distributed in TPTP v4.0.0 were used.
- The CASC framework leverages the TPTP software infrastructure, and thus relied on the new TPTP THF infrastructure developed in the project.
- All of the higher-order theorem proving systems associated with the project - LEO-II, TPS, and IsabelleP - entered the competition.

The CASC-22 web site provides access to all systems and competition resources: `http://www.tptp.org/CASC/22/`. In particular, the THF division results are provided, showing that TPS performed the best, and that all the systems had robust performance on the THF problems. This public demonstration of capability is an important step towards increased usage of higher-order logic and higher-order ATP systems in applications. From 2010 the THF division will be a regular competition division, and CASC will thus be extended to be the world championship for higher-order ATP systems. The THF division is expected to provide insight and stimulus for the development of higher-order ATP systems.

A useful side-effect was derived from Dr. Sutcliffe's work on the THF project at Saarland University - Dr. Christoph Weidenbach (Max-Planck-Institut für Informatik) arranged for the use of the institute's computer cluster for CASC. This is expected to be an ongoing arrangement, thus focussing support for the competition at Saarland University.

3 Impact

3.1 Scientific and Technical

The field of higher-order automated theorem proving is currently experiencing a renaissance. Recently substantial progress has been made with respect to the improvement of existing provers (e.g., TPS and LEO-II) and the development of new provers and model generators (e.g., the new systems of the Isabelle team, and of the group of Prof. Gert Smolka at Saarland University). The THFTPTP project, and in particular the development of the new higher-order TPTP THF infrastructure, has consolidated and strongly fostered these developments. Not only comparison and competition between these systems is now facilitated by the new TPTP THF infrastructure, but also cooperation and exchange of information.

New application domains have been explored, and experiments confirm that higher-order theorem proving systems can solve relevant problems in a range application domains. These are problems that are either not in the scope of less expressive logics, or cannot be solved effectively in these logics. Amongst the new application domains are non-classical logics such as quantified multi-modal logics, and logics for security. Since these logics are simple fragments of classical higher-order logic, the higher-order automated theorem provers can generally be used to solve problems expressed in these logics [**?**, **?**]. Note that there are currently only very few special purpose reasoners available for quantified and multi-modal logics.

Interestingly, the new TPTP THF infrastructure not only supports comparison amongst proof higher-order provers, but also between higher-order and first-order provers. In the future this will be extended to support comparisons between higher-order provers and specialist provers for modal logics.

The new TPTP THF infrastructure supports the uniform integration of higher-order automated theorem provers within interactive or semi-automated proof assistants. These systems are increasingly employed in semi-industrial contexts, but the human interaction costs are still very high. The TPTP THF infrastructure supports the uniform exploitation of higher-order automated provers within proof assistants. Currently a PhD student of Prof. Larry Paulson at Cambridge University is investigating whether an integration of LEO-II into Isabelle/HOL is possible based on the proof representation format developed in the THFTPTP project. The aim is not only to call LEO-II from within Isabelle/HOL, but also to map LEO-II proofs to Isabelle/HOL and hence verify them.

As part of the preparation work for CASC-22 four established proof modes of the TPS system were shown to be unsound. Similarly for LEO-II, the TPTP THF infrastructure helped in the detection of two sources of unsoundness. In all these cases the model generators of the Munich Isabelle/HOL group, IsabelleN and IsabelleM, played a crucial role. Thus, the new TPTP THF infrastructure has already had a significant impact on detecting unsoundness in existing theorem provers, and it will certainly be a very useful resource for debugging novel and existing higher-order theorem provers in the future.

The CASC demonstration competition stimulated interest in the area and several researchers have announced that they will try to contribute further THF compliant reasoners within the next year. The list of researchers includes:

- C. Benzmüller, Dr. Claus-Peter Wirth: A new LEO-III prover is planned that will include more effective methods for treating induction and arithmetic.
- Dr. Chad Brown, Prof. Gert Smolka: They are working on tableau based provers for higher-order logic, with a focus on decidable fragments and automation of the description operator.
- Prof. Koen Claessen: He has expressed a plan in developing a higher-order model generator.
- Dr. Russell Wallace: He is working on higher-order prover based on combinatory logic.

3.2 Further Research

Follow-up project proposals that were strongly influenced by the THFTPTP project:

- C. Benzmüller, LEO-II im Ontologieschliessen. Forschungsstipendium proposal to the Deutsche Forschungsgemeinschaft (DFG). The proposal was granted in June 2009. In this project Benzmüller is collaborating with Articulate Software, USA, applying the higher-order automated theorem prover LEO-II to reasoning in ontologies.
- C. Benzmüller, C.P. Wirth, Effective Higher-Order Automated Theorem Proving with integrated Descente Infinie and Presburger Arithmetic LEO-III. Grant proposal submitted to the Deutsche Forschungsgemeinschaft (DFG).
- S. Autexier, EU ERC Starting Grant. Autexier has proposed a generic reasoning and proof representation framework based on assertion level proofs. This framework is planned to be TPTP THF compliant.

The combined first-order and higher-order TPTP infrastructure now supports experimentation with different logic encodings. For example, it supports the comparison of provers for first-order and higher-order logics for the same abstract problems but different logic level formalizations. Collaborative links have been established with Jens Otten's new project at Postdam University, which aims to build a problem library for multi-modal logics. A future goal is to also facilitate comparison between special purpose provers for quantified multi-modal logics and our higher-order theorem provers.

Interesting future research includes the development of an online intreface for reasoning with multi-modal logics, linked to the existing SystemOnTPTP interface. This will include automated conversion of multimodal logic problems to THF form, and the subsequent application of a higher-order ATP system to solve the problems. This interface could subsequently be extended to also support intuitionistic logics and logics for security (e.g. access control logics). Like multimodal logics, the latter logics are also fragments of THF [3].

Future work includes the development and implementation of a powerful representation, verification, and explanation framework for THF proofs. A suitable, uniform target calculus is envisaged, to which the proof outputs of the different THF provers can be translated. The THF provers will need to adapt their proof output directly to this new format, or transformation tools will have to be provided. The benefits of such a new format are obvious: (i) client systems that embed higher-order ATP system will need to support only this single format and (ii) tools for extracting information from THF proofs, for verifying them, or for explaining them, e.g., to humans, will have to support only this single format. Dr. Serge Autexier at DFKI Bremen has recently proposed a research project in this direction.

In THF0 only simple type theory is supported, and type declarations are provided. Type reconstruction is therefore not an issue. Type checking is handled with the type checking tool based on Twelf. In the richer extensions of the THF0 language, such as the THF and THFX languages, polymorphic and dependent types are allowed. Increasingly advanced type reconstruction and type checking tools will thus be required. The current type checking tool based on Twelf will be extended in our ongoing collaboration with Dr. Carsten Schürmann at IT University Copenhagen and Dr. Florian Rabe at Jacobs International University in Bremen.

Affirmation

The work described in this report was carried out in the THFTPTP Project – An Infrastructure for Typed Higher-order Form Automated Theorem Proving, as a Marie Curie International Incoming Fellowship, Grant Agreement PIIF-GA-2008-219982.

Dr. Christoph Benzmüller (Host)
Saarland University

Prof. Dr. Geoff Sutcliffe (Researcher)
University of Miami

References

[1] C. Benzmüller. Automating Access Control Logics in Simple Type Theory with LEO-II. Technical Report SEKI Technical Report SR-2008-01, Universität des Saarlandes, Saarbrücken, Germany, 2008.

[2] C. Benzmüller. A Note on LEO-II and the Basic Fragment of Simple Type Theory. *AAR Newsletter*, 84, 2009.

[3] C. Benzmüller. Automating Access Control Logic in Simple Type Theory via LEO-II. In *Proceedings of the 24th IFIP International Information Security Conference*, IFIP Advances in Information and Communication Technology, page To appear. Springer-Verlag, 2009.

[4] C. Benzmüller. Automating Quantified Multimodal Logic in Simple Type Theory - A Case Study. Technical Report SEKI Working Paper SWP-2009-02, Universität des Saarlandes, Saarbrücken, Germany, 2009.

[5] C. Benzmüller and L. Paulson. Exploring Properties of Normal Multimodal Logics in Simple Type Theory with LEO-II. In C. Benzmüller, C. E. Brown, J. Siekmann, and R. Statman, editors, *Reasoning in Simple Type Theory: Festschrift in Honour of Peter B. Andrews on his 70th Birthday*, number 17 in Studies in Logic, Mathematical Logic and Foundations. College Publications, 2009.

[6] C. Benzmüller and L. Paulson. Exploring Properties of Propositional Normal Multimodal Logics and Propositional Intuitionistic Logics in Simple Type Theory. *Journal of Symbolic Logic*, 2010.

[7] C. Benzmüller, L. Paulson, F. Theiss, and A. Fietzke. LEO-II - A Cooperative Automatic Theorem Prover for Higher-Order Logic. In P. Baumgartner, A. Armando, and D. Gilles, editors, *Proceedings of the 4th International Joint Conference on Automated Reasoning*, number 5195 in Lecture Notes in Artificial Intelligence, pages 162–170. Springer-Verlag, 2008.

[8] C. Benzmüller, F. Rabe, C. Schürmann, and G. Sutcliffe. Evaluation of Systems for Higher-order Logic. Invited paper at the IJCAR'08 Workshop on Practical Aspects of Automated Reasoning, 2008.

[9] C. Benzmüller, F. Rabe, and G. Sutcliffe. THF0 - The Core TPTP Language for Classical Higher-Order Logic. In P. Baumgartner, A. Armando, and D. Gilles, editors, *Proceedings of the 4th International Joint Conference on Automated Reasoning*, number 5195 in Lecture Notes in Artificial Intelligence, pages 491–506. Springer-Verlag, 2008.

[10] C. E. Brown. Dependently Typed Set Theory. Technical Report SWP-2006-03, Saarland University, 2006.

[11] C. E. Brown. M-Set Models. In C. Benzmüller, C. E. Brown, J. Siekmann, and R. Statman, editors, *Reasoning in Simple Type Theory: Festschrift in Honour of Peter B. Andrews on his 70th Birthday*, number 17 in Studies in Logic, Mathematical Logic and Foundations. College Publications, 2009.

[12] D. Garg. Principal-Centric Reasoning in Constructive Authorization Logic. Technical Report CMU-CS-09-120, School of Computer Science, Carnegie Mellon University, 2009.

[13] D. Garg and M. Abadi. A Modal Deconstruction of Access Control Logics. In R. Amadio, editor, *Proceedings of the 11th International Conference on the Foundations of Software Science and Computational Structures*, number 4962 in Lecture Notes in Computer Science, pages 216–230, 2008.

[14] E. Landau. *Grundlagen der Analysis*. Akademische Verlagsgesellschaft M.B.H, 1930.

[15] T. Raths, J. Otten, and C. Kreitz. The ILTP Problem Library for Intuitionistic Logic - Release v1.1. *Journal of Automated Reasoning*, 38(1-2):261–271, 2007.

[16] G. Sutcliffe. The SZS Ontologies for Automated Reasoning Software. In G. Sutcliffe, P. Rudnicki, R. Schmidt, B. Konev, and S. Schulz, editors, *Proceedings of the LPAR Workshops: Knowledge Exchange: Automated Provers and Proof Assistants, and The 7th International Workshop on the Implementation of Logics*, number 418 in CEUR Workshop Proceedings, pages 38–49, 2008.

[17] G. Sutcliffe and C. Benzmüller. Practical Progress in the Development of Automated Theorem Proving for Higher-order Logic. *Association for Automated Reasoning Newsletter*, (83), 2009.

[18] G. Sutcliffe, C. Benzmüller, C. E. Brown, and F. Theiss. Progress in the Development of Automated Theorem Proving for Higher-order Logic. In R. Schmidt, editor, *Proceedings of the 22nd International Conference on Automated Deduction*, number 5663 in Lecture Notes in Artificial Intelligence, pages 116–130. Springer-Verlag, 2009.

[19] G. Sutcliffe and C. Suttner. The State of CASC. *AI Communications*, 19(1):35–48, 2006.

[20] L.S. van Benthem Jutting. *Checking Landau's "Grundlagen" in the AUTOMATH System*. PhD thesis,

Eindhoven University, Eindhoven, The Netherlands, 1979.