The CADE ATP System Competition — CASC

Geoff Sutcliffe

The CADE ATP System Competition (CASC) is an annual evaluation of fully automatic automated theorem-proving (ATP) systems for classical logic — the world championship for such systems. CASC provides a public evaluation of the relative capabilities of ATP systems, and aims to stimulate ATP research toward the development of more powerful ATP systems. Over the years CASC has been a catalyst for impressive improvements in ATP.
Competition Reports

...
using code trees and advanced structures for representing logical data, is repeatedly tested and tuned on the TPTP problem library, and has special modes for the various divisions of CASC. Technical information about these systems, and the techniques they employ, can be found on the individual CASC web pages.

The design and organization of CASC have evolved over the years to a sophisticated state. Decisions made for CASC (alongside the TPTP) have had an influence on the directions of development in ATP. It is interesting to look back on some of the key decisions that have helped bring the competition to its current state.

CASC-13, 1996 — the first CASC — stimulated research toward robust, fully automatic systems that take only logical formulae as input. It increased the visibility of systems and developers, and rewarded implementation efforts. CASC-14, 1997, introduced the SAT division, stimulating the development of model-finding systems for CNF. CASC-15, 1998, introduced the FOF division, starting the slow demise of CNF to becoming just the assembly language of ATP. At CASC-16 in 1999, changes to the problem-selection process motivated the development of techniques for automatic tuning of ATP systems’ search parameters. CASC-JC, 2001, introduced ranking based on proof output, starting the trend toward ATP systems that efficiently output proofs and models. CASC-JC also introduced the EPR division, stimulating the development of specialized techniques for this important subclass of problems. CASC-20, 2005, required systems to develop built-in equality reasoning, by removing the equality axioms from all TPTP problems. At CASC-J3, in 2006, the FOF division was promoted as the most important, stimulating development of ATP systems for full first-order logic. CASC-21, 2007, introduced the FNT division, further stimulating the development of model-finding systems. CASC-J4, 2008, introduced the LTB division, stimulating the development of techniques for automatically dealing with very large axiom sets. CASC-J5, 2010, introduced the THF division, stimulating development of ATP systems for higher-order logic. CASC-23, 2011, introduced the TFA division, stimulating development of ATP systems for full first-order logic with arithmetic. At CASC-J6, in 2012, Prover9 replaced Otter as the fixed-point in the FOF division, demonstrating the progress in ATP. CASC-24, 2013, removed the CNF division, confirming the demise of CNF. CASC-J7, 2014, required use of the SBS ontology, so the ATP systems unambiguously report what they have established about the problem. CASC-25, 2015, introduced the THN and TFN divisions, stimulating development of model finding for the THF and TFA logics.

Over the years TPTP and CASC have increasingly been used as a conduit for ATP users to provide samples of their problems to ATP system developers.

Users’ problems that are contributed to TPTP are eligible for use in CASC. The problems are then exposed to ATP system developers, who improve their systems’ performances on the problems, in order to perform well in CASC. This completes a cycle that provides the users with more effective tools for solving their problems.

Notes
1. Christian Suttner was a CASC organizer for the first 10 CASCs, and various other people have contributed to the running of selected CASC editions.
2. www.tptp.org/CASC.
3. Effectively propositional means that the problem is known to be reducible to a propositional problem, e.g., a CNF problem that has no functions with arity greater than zero.

References

Geoff Sutcliffe is a professor and chair of the Department of Computer Science at the University of Miami. His research is in the area of automated reasoning, particularly in the evaluation and effective use of automated reasoning systems. His most prominent achievements are the first ever development of a heterogeneous parallel deduction system, leading to the development of the SSCPA automated reasoning system; the development and ongoing maintenance of the TPTP problem library, which is now the de facto standard for testing classical logic automated reasoning systems; the development and ongoing organization of the CADE ATP System Competition — the world championship for classical logic automated reasoning systems; and the specification of the TPTP language standards for automated reasoning tools.