EU Project THFTPTP: An Infrastructure for HOL-ATP

Results of the EU Project THFTPTP

- Collaboration with Geoff Sutcliffe, Chad Brown and others
- Results
  - THF0 syntax for HOL
  - Online access to provers
  - Library with example problems (e.g. entire TPS library) and results
  - Ontology and syntax for proof results
  - International CASC competition for HOL-ATP
  - Various tools

Improved availability and robustness of HOL-ATPs: TPS, LEO-II, Isabelle, Satallax, Refute, Nitpick, agsyHOL   http://www.tptp.org/cgi-bin/SystemOnTPTP

[SutcliffeBenzmüller, J. Formalized Reasoning, 2010]

[BenzmüllerRabeSutcliffe, IJCAR, 2008]

Automated Reasoners for HOL

TPS ... ———————————— (Peter Andrews) ——————— ?

LEO-I/LEO-II (myself)

Isabelle (Nipkow/Paulson/Blanchette)

Satallax (Brown)

Nitpick (Blanchette)

agsyHOL (Lindblatt)

- all accept TPTP THF Syntax [SutcliffeBenzmüller, J.Form.Reas, 2009]
  - can be called remotely via SystemOnTPTP at Miami
  - they significantly gained in strength over the last years
  - they can be bundled into a combined prover HOL-P

> Exploit HOL with Henkin semantics as metalogic
> Automate other logics (& combinations) via semantic embeddings
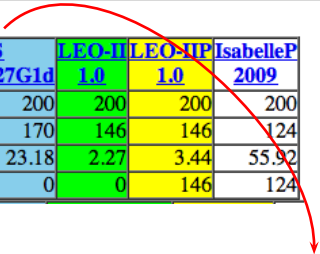> — HOL-P becomes a Universal Reasoner —

CASC Competitions in THF (HOL) Category since 2009

2009

THF	TPS 3.20080227G1d	LEO-II 1.0	LEO-IIP 1.0	IsabelleP 2009
Attempted	200	200	200	200
Solved	170	146	146	124
Av. Time	23.18	2.27	3.44	55.92
Solutions	0	0	146	124

CASC Competitions in THF (HOL) Category since 2009

2009

THF	TPS 3.20080227G1d	LEO-II 1.0	LEO-IIP 1.0	IsabelleP 2009
Attempted	200	200	200	200
Solved	170	146	146	124
Av. Time	23.18	2.27	3.44	55.92
Solutions	0	0	146	124

2010

THF	LEO-II 1.2	Satallax 1.4	IsabelleP 2009-2	TPS 3.20080227G1d
Solved	125/200	120/200	101/200	80/200
Av. CPU Time	16.65	55.24	100.75	36.15
Solutions	125/200	120/200	0/200	0/200

LEO-II 1.2 solved 56% more than previous winner

CASC Competitions in THF (HOL) Category since 2009

2009

THF	TPS 3.20080227G1d	LEO-II 1.0	LEO-IIP 1.0	IsabelleP 2009
Attempted	200	200	200	200
Solved	170	146	146	124
Av. Time	23.18	2.27	3.44	55.92
Solutions	0	0	146	124

2010

THF	LEO-II 1.2	Satallax 1.4	IsabelleP 2009-2	TPS 3.20080227G1d
Solved	125/200	120/200	101/200	80/200
Av. CPU Time	16.65	55.24	100.75	36.15
Solutions	125/200	120/200	0/200	0/200

LEO-II 1.2 solved 56% more than previous winner

2011

THF/300	Satallax 2.1	LEO-II 1.2.8	LEO-II 1.2	Isabelle 2011	TPS 3.110228S1a
Solved	246/300	208/300	204/300	201/300	190/300
Av. CPU Time	12.04	8.97	4.95	36.55	18.69

Satallax 2.1 solved 21% more than previous winner

CASC Competitions in THF (HOL) Category since 2009

2010

THF	LEO-II 1.2	Satallax 1.4	IsabelleP 2009-2	TPS 3.20080227G1d
Solved	125/200	120/200	101/200	80/200
Av. CPU Time	16.65	55.24	100.75	36.15
Solutions	125/200	120/200	0/200	0/200

LEO-II 1.2 solved 56% more than previous winner

2011

THF /300	Satallax 2.1	LEO-II 1.2.8	LEO-II 1.2	Isabelle 2011	TPS 3.110228S1a
Solved	246/300	208/300	204/300	201/300	190/300
Av. CPU Time	12.04	8.97	4.95	36.55	18.69

Satallax 2.1 solved 21% more than previous winner

2012

Higher-order Theorems	Isabelle-H 2012	Isabelle 2012	Satallax 2.4	Satallax 2.1	LEO-II 1.4.0	TPS 3.120601S1b
Solved/200	166/200	135/200	132/200	123/200	81/200	66/200
Av. CPU Time	88.44	70.13	16.20	19.57	11.38	25.23

Isabelle-HOT solved 35% more than previous winner

CASC Competitions in THF (HOL) Category since 2009

2012

Higher-order Theorems	Isabelle-H 2012	Isabelle 2012	Satallax 2.4	*Satallax* 2.1	LEO-II 1.4.0	TPS 3.120601S1b
Solved/200	166/200	135/200	132/200	123/200	81/200	66/200
Av. CPU Time	88.44	70.13	16.20	19.57	11.38	25.23

CASC Competitions in THF (HOL) Category since 2009

2012

Higher-order Theorems	Isabelle-H 2012	Isabelle 2012	Satallax 2.4	*Satallax* 2.1	LEO-II 1.4.0	TPS 3.120601S1b
Solved/200	166/200	135/200	132/200	123/200	81/200	66/200
Av. CPU Time	88.44	70.13	16.20	19.57	11.38	25.23

LEO-II cooperates with FOL prover E

2012

Higher-order Theorems	Isabelle-H 2012	Isabelle 2012	Satallax 2.4	*Satallax* 2.1	LEO-II 1.4.0	TPS 3.120601S1b
Solved/200	166/200	135/200	132/200	123/200	81/200	66/200
Av. CPU Time	88.44	70.13	16.20	19.57	11.38	25.23

Satallax cooperates with SAT solver Minisat

2012

Higher-order Theorems	Isabelle-H 2012	Isabelle 2012	Satallax 2.4	Satallax 2.1	LEO-II 1.4.0	TPS 3.120601S1b
Solved/200	166/200	135/200	132/200	123/200	81/200	66/200
Av. CPU Time	88.44	70.13	16.20	19.57	11.38	25.23

Isabelle-HOT cooperates with various FOL provers (sledgehammer) and SMT solvers (smt) and even with LEO-II and Satallax