The ILTP Library: Benchmarking Automated Theorem Provers for Intuitionistic Logic

Thomas Raths Jens Otten Christoph Kreitz

Institut für Informatik, University of Potsdam
August-Bebel-Str. 89, 14482 Potsdam-Babelsberg, Germany
{raths,jeotten,kreitz}@cs.uni-potsdam.de

Abstract. The Intuitionistic Logic Theorem Proving (ILTP) library provides a platfom for testing and benchmarking theorem provers for first-order intuitionistic logic. It includes a collection of benchmark problems in a standardized syntax and performance results obtained by a comprehensive test of currently available intuitionistic theorem proving systems. These results also provide information about the status and the difficulty rating of the benchmark problems.

1 Introduction

Benchmarking automated theorem proving (ATP) systems using standardized problem sets is a well-established method for measuring their performance. The TPTP library [14, 15] is the largest collection of problems (currently more than 7000) for testing and benchmarking ATP systems for classical logic. Other problem libraries for, e.g., termination problems [8] or induction problems [6, 4] have been developed as well.

Unfortunately the availability of such libraries for non-classical logics is very limited. For intuitionistic logic several small collections of formulae have been published and used for testing ATP systems. One of the first collections of first-order formulae has been compiled in [12] where it has been used for testing the intuitionistic ATP system ft. The same collection has also been used for benchmarking the intuitionistic theorem provers in [16, 9]. A second collection of first-order formulae has been used in [13] to test and benchmark the intuitionistic ATP system JProver, which has been integrated into the constructive interactive proof assistants NuPRL [3] and Coq [2].

Another collection of propositional formulae has been compiled by Dyckhoff in [5]. It introduces six classes of scalable formulae following the methodology of the Logics Workbench [1]. The advantage of this approach is the possibility to study the time complexity behaviour of an ATP system on a specific generic formula as its size increases. But in order to achieve more meaningful benchmark results the number of generic formulae would have to be increased significantly. Most of the formulae in the collection have a rather syntactical nature, often specifically designed with the presence (or absence) of a specific search strategy in mind. To provide a better view of the usefulness of intuitionistic ATP systems on problems arising in practice, like in program synthesis [3], a benchmark collection should cover a broader range of more realistic problems. These kind of problems are typically presented in a first-order logic (as already mentioned in [5]).

The ILTP library was developed for even that purpose In the following we will describe the content of ILTP library, which contains two major problem sets, a database of currently available intuitionistic ATP systems with performance results, and some benchmark tools. We will also present comprehensive tests with existing intuitionistic ATP systems yield information about the intuitionistic status and difficulty rating of the problems in the ILTP library.

2 The Content of the ILTP Library

The ILTP library contains two main set of problems: one problem set is taken from the TPTP library and the other is taken from problem sets which have been used previously for testing and benchmarking intuitionistic ATP systems.

2.1 The TPTP Problem Set

Whereas the semantics of classical and intuitionistic logic differs, they share the same syntax. This allows in principal the use of classical benchmark libraries like the TPTP library [14, 15] for benchmarking intuitionistic ATP systems as well. But some of the equivalences, like De Morgan's law $\neg(A \wedge B) \Leftrightarrow (\neg A \vee \neg B)$, used to translate formulae into clausal (i.e. disjunctive or conjunctive normal) form are not valid in intuitionistic logic. It is also not possible to prove a formula by the regularly applied practice of refuting its negation, since $(\neg A \Rightarrow \bot) \Rightarrow A$, i.e. $\neg\neg A \Rightarrow A$, is not valid either.

Starting mainly as a library of first-order formulae in clausal form, today the TPTP library contains a large number of formulae in non-clausal form as well. The TPTP library version 2.7.0 contains 1745 problems in non-clausal form, so called "first-order formulae" (FOF). Of these formulae 408 are classically invalid. Since every intuitionistically valid formula needs to be classically valid as well, it is straightforward to refute these formulae with a classical ATP system. Therefore we will focus on the remaining 1337 formulae whose classical status is either valid or unknown.

These 1337 formulae form the first part, the TPTP problem set, of the ILTP library. The status (i.e. Theorem, Non-Theorem, Unknown) and the difficulty rating of the problems have been adapted to the intuitionistic case (see Section 3) and are provided separately.

2.2 The ILTP Problem Set

The second part of the ILTP library contains formulae from the following three collections of benchmark problems.

The first collection contains 39 intuitionistically valid first-order formulae originally used to test the intuitionistic ATP system ft [14]. Five of the problems are already part of the TPTP problem set mentioned above. These are problems ft3.1 to ft3.5 which are identical with Pelletier's problems no. 39 to 43.

The second collection contains 36 propositional formulae from Dyckhoffs' benchmark collection in [5]. From each of the six problem classes three (intuitionistically) valid and three invalid formulae have been included. These six formula instances have been chosen according to their difficulty relative to current intuitionistic ATP systems.

The third collection contains 33 propositional and first-order formulae from the problem set used to test the intuitionistic ATP system JProver [13]. The type information which has been used to test JProver within the NuPRL environment has been removed. Three poblems have been left out because they are already classically invalid or cannot be represented in pure first-order logic.

Altogether this second set of problems, which we call the ILTP problem set, includes 108 formulae. Their syntax have been standardized and adapted to the TPTP input format. Each problem file has been given a header with useful information, like references, like done in the TPTP library. The intuitionistic status and difficulty rating (see Section 3) has been included as well.

2.3 Prover Database and Tools

In addition to the two problems sets, the ILTP library contains a small database with information about published intuitionistic ATP systems. For each prover we provide some basic information (like author, homepage, short description, references) and a test run on two example formulae. A summary and a detailed list of the performance results on running each system on the problem in the ILTP library are given as well.

We also provide so-called format files, which can be used to convert the problems in the ILTP library into the input syntax of the ATP systems listed in the prover database. These format files are used together with the TPTP2X utility which is part of the TPTP library.

3 Rating the Problems in the ILTP Library

In the TPTP library the difficulty of every problem is rated according to the performance of current state-of-the-art ATP systems. It expresses the ratio of systems which can solve a problem. For example a rating of 0.0 indicates that every state-of-the-art prover can solve the problem, a rating of 0.5 indicates that half of the systems were able to solve it, and a problem with rating 1.0 has not been solved by any ATP system.

We adapt this notation to the problems in the ILTP library. To this end we need to specify a set of intuitionistic state-of-the-art ATP systems. We performed comprehensive tests of currently available systems on the problems in the ILTP library and analysed the performance results [11]. We have selected the following five intuitionistic ATP systems which solve at least one problem which no other system was able to solve: the first-order systems ft (C-version) [12], JProver [13], ileanTAP[9], ileanCoP[10], and the propositional system STRIP [7].

Each problem is assigned its status. The status can be Theorem, Non-Theorem or Unknown. We did not perform any theoretical investigations into the intuitionistic validity of the formulae in the TPTP problem set. We mark the status of a problem as Theorem or Non-Theorem if any ATP system was able to show that the given problem is valid or invalid, respectively. All other TPTP problems were given the status Unknown.

3.1 TPTP Problem Set

Figure 1 shows a summary of the rating and status information of the TPTP problem set. The rating and status information refers to intuitionistic logic. Only the last line shows the (original TPTP) classical rating of the problem set.

Rating	0.0	0.01–0.25	0.26–0.50	0.51–0.75	0.76–0.99	1.0	Σ
Theorem	74	21	28	97	0	0	220
Non-Theorem	2	0	5	45	1	0	53
Unknown	0	0	0	0	0	1064	1064
Classical	286	245	102	256	265	183	1337

Domain		AGT	ALG	COM	GEO	LCL	MGT	NLP	PUZ	SET	SWV	SYN
Theorem		14	7	3	7	1	25	11	2	75	1	74
Non-Theorem		0	1	0	0	2	0	0	0	0	0	50
Unknown		38	137	0	65	0	42	11	2	244	1	92
intuit.	0.0	0	0	0	0	1	5	3	0	14	1	52
	>0.0	52	145	3	72	2	62	19	4	305	1	164
classic.	0.0	43	6	1	0	3	29	7	3	40	1	123
	>0.0	9	139	2	72	0	38	15	1	279	1	93

Fig. 1. Rating of the TPTP Problem Set

From the 1337 problems in the TPTP set 220 have been proven intuitionistically valid, 53 intuitionistically invalid.

3.2 ILTP Problem Set

Figure 2 shows a summary of the rating and status information of the ILTP problem set. Again the rating and status information refers to intuitionistic logic.

From the 108 problems in the ILTP set 90 are known to be intuitionistically valid, 18 invalid.

Rating	0.0	0.01–0.25	0.26–0.50	0.51–0.75	0.76–0.99	1.0	Σ
Theorem	59	11	10	8	1	1	90
Non-Theorem	0	0	5	9	3	1	18
Propositional	14	3	8	17	4	2	48
First-order	45	8	7	0	0	0	60

Fig. 2. Rating of the ILTP Problem Set

4 Conclusion

Like the TPTP library for classical logic, the main motivation for the ILTP library is to put the testing and evaluation of intuitionistic ATP systems onto a firm basis. This will help to ensure that published results reflect the actual performance of an ATP system and make meaningful system evaluations and comparisons possible. We hope that the existence of such a library is fruitful for the development of novel, more efficient calculi and implementations for intuitionistic first-order logic, which — compared to classical logic — is still in its infancy. We have mainly focused on the first-order logic which is practically more relevant than the propositional fragment. Future work includes adding more formulae which occur during the practical use of interactive proof systems (like [3]). Extending the library to other non-classical logics like first-order modal logics or fragments of linear logic is under consideration as well.

The ILTP library is available at `http://www.cs.uni-potsdam.de/ti/iltp`.

References

1. P. Balsiger, A. Heuerding, S. Schwendimann. Logics Workbench 1.0. *7th TABLEAUX Conference*, LNCS 1397, Springer, pp. 35–37, 1998.
2. Y. Bertot, P. Castéran. *Interactive Theorem Proving and Program Development*. Texts in Theoretical Computer Science, Springer, 2004.
3. R. L. Constable et. al. *Implementing Mathematics with the NuPRL proof development system*. Prentice Hall, 1986.
4. L. Dennis. Induction challenge problems. 2005. http://www.cs.nott.ac.uk/~lad/research/challenges/
5. R. Dyckhoff. Some benchmark formulae for intuitionistic propositional logic. University of St Andrews, 1997. http://www.dcs.st-and.ac.uk/~rd/logic/marks.html
6. I. Green. Dream corpus of inductive conjectures. 2005. http://dream.dai.ed.ac.uk/dc/lib.html
7. D. Larchey-Wendling, D. Méry, D. Galmiche. STRIP: Structural Sharing for Efficient Proof-Search. *IJCAR-2001*, LNAI 2083, pp. 696–700, Springer, 2001.
8. C. March, A. Rubio, H. Zantema. The termination problems database. 2005. http://www.lri.fr/~marche/tpdb/.
9. J. Otten. ileanTAP: An intuitionistic theorem prover. *6th TABLEAUX Conference*, LNAI 1227, pp. 307–312, Springer, 1997.
10. J. Otten. Clausal connection-based theorem proving in intuitionistic first-order logic. Technical Report, University of Potsdam, 2005. Submitted.

11. T. Raths. Testen von Theorembeweisern. Technical Report, University of Potsdam, 2005.
12. D. Sahlin, T. Franzen, S. Haridi. An Intuitionistic Predicate Logic Theorem Prover. *Journal of Logic and Computation*, 2:619–656, 1992.
13. S. Schmitt et al. JProver: Integrating connection-based theorem proving into interactive proof assistants. *IJCAR-2001*, LNAI 2083, pp. 421–426, Springer, 2001.
14. G. Sutcliffe, C. Suttner, T. Yemenis. The TPTP problem library. *12th CADE*, LNAI814, pp. 252–266, Springer, 1994. http://www.cs.miami.edu/∼tptp
15. G. Sutcliffe, C. Suttner. The TPTP problem library - CNF release v1.2.1. *Journal of Automated Reasoning*, 21: 177–203, 1998.
16. T. Tammet. A Resolution Theorem Prover for Intuitionistic Logic. *13th CADE*, LNAI 1104, pp. 2–16, Springer, 1996.