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 standardised syntax and performance results obtained by a comprehensive test of currently available intuitionistic theorem proving systems. These results are used to provide information about the status and the difficulty rating of the benchmark problems.

1 Introduction

Benchmarking automated theorem proving (ATP) systems using standardised problem sets is a well-established method for measuring their performance. The TPTP library [10] is the largest collection of problems (currently more than 7000 formulas) for testing and benchmarking ATP systems for classical logic. Other problem libraries for, e.g., termination and induction problems have been developed as well.[1]

Unfortunately the availability of such libraries for non-classical logics is very limited. For intuitionistic logic several small collections of formulas have been published and used for testing ATP systems. Sahlin et al. [8] compiled one of the first collections of first-order formulas for testing their intuitionistic ATP system ft. The same collection was also used for benchmarking other intuitionistic theorem provers [11, 5]. A second collection of first-order formulas was used to test the intuitionistic ATP system JProver [9], which has been integrated into the constructive interactive proof assistants NuPRL [3] and Coq [2].

Another collection of propositional formulas was compiled by Dyckhoff.[2] It introduces six classes of scalable formulas 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

* The first author's research is sponsored by DARPA under agreement number FA8750-04-2-0216.

[1] For the termination problem library see http://www.lri.fr/~marche/tpdb/, for the induction problem libraries see http://dream.dai.ed.ac.uk/dc/lib.html and http://www.cs.nott.ac.uk/~lad/research/challenges/.

[2] See http://www.dcs.st-and.ac.uk/~rd/logic/marks.html .

as its size increases. But in order to achieve more meaningful benchmark results the number of generic formulas would have to be increased significantly. Most of the formulas 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 Dyckhoff's benchmark collection).

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

2 The Content of the ILTP Library

The ILTP library contains two main set of problems: the first one is taken from the TPTP library and the second one from three problem collections, 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 principle the use of classical benchmark libraries like the TPTP library [10] for benchmarking intuitionistic ATP systems as well. Starting mainly as a library of first-order formulas in clausal form, today the TPTP library contains a large number of formulas in non-clausal form as well. Problems in clausal (i.e. disjunctive or conjunctive normal) form are intuitionistically invalid and therefore useless for intuitionistic reasoning. Furthermore, the conversion of formulas to clausal form does not preserve intuitionistic validity, because it involves (intuitionistically invalid) laws like $\neg(A \wedge B) \Rightarrow (\neg A \vee \neg B)$ and $\neg\neg A \Rightarrow A$. Adding double negation to classically valid formulas in order to generate intuitionistically valid formulas is of less interest as well since the resulting problems are just encodings of the classical ones.

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

These 1337 formulas 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, which we call the ILTP problem set, contains 108 formulas from three benchmark collections. Their syntax was standardised and adapted to the TPTP input format. Each problem file was given a header with useful information, like references, as done in the TPTP library. The intuitionistic status and difficulty rating (see Section 3) was included as well.

The first collection contains 39 intuitionistically valid first-order formulas originally used to test the intuitionistic ATP system ft [8]. Five of the problems are already part of the TPTP problem set and therefore excluded. 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 formulas from Dyckhoff's benchmark collection. From each of the six problem classes three (intuitionistically) valid and three invalid formulas are 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 formulas from the problem set used to test the intuitionistic ATP system JProver [9]. The type information, which was used to test JProver within the NuPRL environment, is removed. Three problems are left out because they are already classically invalid or cannot be represented in pure first-order logic.

2.3 Tools and Prover Database

In addition to the two problems sets, we 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. The ILTP library also 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 formulas. A summary and a detailed list of the performance results on running each system on the problems in the ILTP library are given as well.

3 Rating the Difficulty of Problems in the ILTP Library

In the TPTP library the difficulty of every problem is rated according to the performance of current (classical) 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 was not 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 all currently available systems on the problems

in the ILTP library and analysed the performance results [7]. We have selected four first-order and one purely propositional ATP system, which solved the highest number of problems: the first-order systems ft (C-version) [8], JProver [9], ileanTAP[5], ileanCoP[6], and the propositional system STRIP [4].

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 formulas 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 The TPTP Problem Set

Table 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.

Table 1. Rating of the TPTP 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

The lower part of Table 1 contains information with respect to the TPTP problem domain (e.g. SET contains problems from set theory). All problems of the domains GRP, HAL and SWC are Unknown (and not included). From the 1337 problems 220 have been proven intuitionistically valid, 53 invalid.

Table 2. Rating of the ILTP 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	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

3.2 The ILTP Problem Set

Table 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 90 are intuitionistically valid, 18 are invalid.

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. It is the first systematic attempt to assemble a benchmark library for intuitionistic theorem provers. 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 expect that such a library will be 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 first-order logic which is practically more relevant than the propositional fragment. Future work includes adding more formulas which occur during the practical use of interactive proof assistants like NuPRL [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.iltp.de`. We welcome submissions of new intuitionistic benchmark problems and intuitionistic ATP systems.

References

1. P. Balsiger, A. Heuerding, S. Schwendimann. Logics workbench 1.0. *7^{th} 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. D. Larchey-Wendling, D. Méry, D. Galmiche. STRIP: Structural sharing for efficient proof-search. *IJCAR-2001*, LNAI 2083, pp. 696–700, Springer, 2001.
5. J. Otten. ileanTAP: An intuitionistic theorem prover. *6^{th} TABLEAUX Conference*, LNAI 1227, pp. 307–312, Springer, 1997.
6. J. Otten. Clausal connection-based theorem proving in intuitionistic first-order logic. *TABLEAUX 2005*, this volume, 2005. (`http://www.leancop.de`)
7. T. Raths. Evaluating intuitionistic automated theorem provers. Technical Report, University of Potsdam, 2005.
8. D. Sahlin, T. Franzen, S. Haridi. An intuitionistic predicate logic theorem prover. *Journal of Logic and Computation*, 2:619–656, 1992.
9. S. Schmitt et al. JProver: Integrating connection-based theorem proving into interactive proof assistants. *IJCAR-2001*, LNAI 2083, pp. 421–426, Springer, 2001.
10. G. Sutcliffe, C. Suttner. The TPTP problem library - CNF release v1.2.1. *Journal of Automated Reasoning*, 21: 177–203, 1998. (`http://www.cs.miami.edu/~tptp`)
11. T. Tammet. A resolution theorem prover for intuitionistic logic. *13^{th} CADE*, LNAI 1104, pp. 2–16, Springer, 1996.