The ILTP Problem Library for Intuitionistic Logic

Release v1.1

Thomas Raths*, Jens Otten and Christoph Kreitz
Institut für Informatik, University of Potsdam, Germany.
e-mail: {traths,jeotten,kreitz}@cs.uni-potsdam.de

Abstract. The Intuitionistic Logic Theorem Proving (ILTP) library provides a platform for testing and benchmarking automated theorem proving (ATP) systems for intuitionistic propositional and first-order logic. It includes about 2800 problems in a standardized syntax from 24 problem domains. For each problem an intuitionistic status and difficulty rating were obtained by running comprehensive tests of currently available intuitionistic ATP systems on all problems in the library. Thus for the first time the testing and evaluation of ATP systems for intuitionistic logic is put onto a firm basis.

Keywords: ILTP, problem library, benchmarking, ATP, intuitionistic logic

1. Introduction

Benchmarking automated theorem proving (ATP) systems using standardized problem sets is a well-established method for measuring their performance. Several such problem libraries have been developed, e.g. the TPTP library for classical first-order logic (Sutcliffe and Suttner, 1998), libraries for (propositional) satisfiability problems, e.g. SATLIB (Hoos and Stützle, 2000), and libraries for termination and induction problems.[1]

The main purpose of such libraries is to put the testing and evaluation of ATP systems onto a firm basis. The practice of using a few specific problems, possibly even tailored towards a particular ATP system, does not result in meaningful data and therefore does not reflect the actual performance of an ATP system. Using a common standardized and comprehensive problem library yields meaningful results and ensures fair system comparisons.

Unfortunately the availability of such libraries for non-classical logics, like intuitionistic or modal first-order logics, is very limited. The aim of the Intuitionistic Logic Theorem Proving (ILTP) library is to close this gap for the intuitionistic first-order and propositional logic. The

* 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/`.

ILTP library fulfils the following main requirements of a benchmark library as described in (Sutcliffe and Suttner, 1998):

- It is easy to discover and obtain and provides guidelines for its use in evaluating ATP systems; the ILTP library is widely available via the Internet (see Subsection 1.1).
- It is well structured, documented and provides statistics about the library as a whole (see Section 2).
- It is easy to use; the problems are provided in an easy to understand format and conversion tools to other known syntax formats are included (see Subsections 2.3 and 2.4).
- It is large enough for statistically significant testing; the current release v1.1 of the ILTP library contains about 2800 problems.
- It contains problems of varying difficulty; the ILTP library includes simple as well as unsolved/open problems (see Subsection 2.2).
- It assigns each problem an unique name and provides status and difficulty rating for each problem; comprehensive tests of existing intuitionistic ATP systems yielded *intuitionistic* status and rating information (see Subsection 2.2).

This paper describes the release v1.1.1 of the ILTP library. It starts by providing information on how to obtain and use the library, followed by an overview of previous benchmark collections and a description of changes made with respect to release v1.0 of the library. Section 2 gives detailed information about the contents of the ILTP library, which is divided into a first-order and a propositional part. It describes the domain structure, the presentation form of the problems and presents statistics about intuitionistic status and difficulty rating. Some details of included tools are given as well. The paper concludes in Section 3 with a summary and an outlook on further research.

1.1. Obtaining and Using the ILTP Library

The ILTP library is available at `http://www.iltp.de`. It consists of a first-order part and a purely propositional part, which can be downloaded separately. The library is structured into four subdirectories:

`Axioms` - contains the axiom files (only in the first-order part).
`Documents` - contains papers and statistic files.
`Problems` - contains a directory for each domain with problem files.

TPTP2X - contains the tptp2X tool and the format files.

The following conditions should be observed when presenting results of ATP systems based on the ILTP library:

- The release number of the ILTP library has to be stated.
- Each problem should be referred to by its unique name.
- No part of the problems may be modified; no reordering of axioms, hypotheses and/or conjectures is allowed. Only the syntax of problems may be changed, e.g., by using the tptp2X tool (see Subsection 2.4).
- The header information of each problem may not be exploited by an ATP system.
- The version of the tested ATP system including all settings must be documented.

It is a good practice to make at least the binary/executable of an ATP system available whenever performance results or statistics based on the ILTP library are given. This makes the verification and validation of the given performance data possible.

1.2. Previous Problem Collections

For intuitionistic logic several small collections of problems have been published and used for testing existing ATP systems. Sahlin et al. (Sahlin et al., 1992) compiled one of the first collections of first-order problems for testing their intuitionistic ATP system ft. The same collection was also used for benchmarking other intuitionistic ATP systems (Tammet, 1996; Otten, 1997). A second collection of first-order problems was used to test the intuitionistic ATP system JProver (Schmitt et al., 2001), which has been integrated into the constructive interactive proof assistants NuPRL (Allen et al., 2000) and Coq (Bertot and Castéran, 2004).

A collection of propositional problems was compiled by Dyckhoff.[2] It introduces six classes of scalable problems following the methodology of the Logics Workbench (Balsiger et al., 1998). The advantage of this approach is the possibility to study the time complexity behaviour of an ATP system on a specific generic problems as its size increases. But in order to achieve more meaningful benchmark results the number

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

of generic problems would have to be increased significantly. Most of the problems 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, 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.

For classical logic the well-known TPTP library (Sutcliffe and Suttner, 1998) provides a large collection of first-order problems — currently more than 8000 — for testing and benchmarking ATP systems for classical logic. 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 for benchmarking intuitionistic ATP systems as well.

The TPTP library started as a library of first-order formulas in clausal form (Sutcliffe and Suttner, 1998) and the majority of problems are still in clausal form. Problems in clausal form, 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 since the resulting problems are just encodings of the classical ones. Today the TPTP library contains a large number of problems in non-clausal form as well. These problems form a main part of the ILTP library.

1.3. A Short History of the ILTP Library

First experimental evaluations of the intuitionistic first-order ATP systems JProver (Schmitt et al., 2001) and ileanTAP (Otten, 1997) on a large set of suitable problems from the TPTP library were conducted in June 2004. Since then the number of problems as well as the number of tested intuitionistic ATP systems was continuously increased. The first public release v1.0 of the ILTP library (Raths et al., 2005) went online in April 2005. Release v1.1 came out in January 2006. These are the major changes and enhancements that were done compared to release v1.0 of the ILTP library:

- Due to the substantial interest in developing ATP systems for intuitionistic *propositional* logic, the library was split up into a propositional part and a first-order part.

- The number of problems in the library has almost doubled from 1445 problems in the first release to 2754 problems. Three new problem domains have been added.

- The library now includes all non-clausal (FOF) problems of the TPTP library release v3.1.0. The first release of the ILTP library was mainly based on release v2.7.0 of the TPTP library. The format files were adapted to the new TPTP syntax.

- The number of intuitionistic ATP systems considered for determining the intuitionistic status and rating information was increased to eight systems chosen from a total of 14 tested systems. The newest version of all systems was considered. The time limit for status and rating evaluations was increased to 600 seconds.

2. Contents of the ILTP Library

The release v1.1.1 of ILTP library includes a total of 2363 abstract problems. Some of these problems have alternative presentations or are generic and allow to scale the actual problem size. This results in a total of 2754 problems. Table I shows the overall statistics of the problems in the library.

Table I. Overall statistics of the ILTP library v1.1.1

Number of problem domains	24	
Number of abstract problems	2363	
Number of generic problems	20	
Total number of problems	2754	(100%)
Number of non-propositional problems	2480	(90%)
Number of propositional problems	274	(10%)
Number of problems with equality	1730	(63%)
Number of pure equality problems	185	(7%)

Table II provides two separate statistics about the first-order and the propositional part of the library. The intuitionistic status and rating is explained in Subsection 2.2.

The 2754 problems are divided into 24 problem domains. The following subsections provide details about the problem domains, the sources they have been collected from, the intuitionistic status and

Table II. Number of problems in the ILTP library v1.1.1

Status	Theorem	Non-Theorem	Unsolved	Open	$\sum$
First-order part	535	96	424	1495	2550
Propositional part	128	109	37	0	274

difficulty rating, the problem naming and presentation norm, and the tools included in the ILTP library.

2.1. The ILTP Domain Structure

The problems of the ILTP library are grouped into 24 problem domains. The major source of problems is the TPTP library (Sutcliffe and Suttner, 1998). Release 3.1.0 of the TPTP library includes 2324 problems in non-clausal form, so-called "first-order form" (FOF). These problems are classified into 21 domains. Each domain is identified by a three-letter mnemonic, which is also component of the naming scheme (see Subsection 2.3). A more detailed description of these domains can be found in (Sutcliffe and Suttner, 1998).

The domains are as follows: AGT (agents), ALG (general algebra), COM (computing theory), CSR (commonsense reasoning), GEO (geometry), GRA (graph theory), GRP (group theory), HAL (homological algebra), KRS (knowledge representation), LCL (logic calculi), MGT (management), MSC (miscellaneous), NLP (natural language processing), NUM (number theory), PLA (planning), PUZ (puzzles), SET (set theory), SWC (software creation), SWV (software verification), SYN (syntactic), and TOP (topology). Statistics about the problems can be found in Table III and IV.

Additionally three more domains with a total of 430 problems are included. These domains are GEJ (constructive geometry), GPJ (non-clausal group theory) and SYJ (intuitionistic syntactic) and contain the following problems.

GEJ – *Constructive Geometry.*
Elementary geometry was formalized constructively according to (von Plato, 1995). The original axiomatization and a shortened one (Dafa, 1997; Dafa, 1998) are used. Some problems are formalized using the reduced axiom set described in (Tammet, 1996).

GPJ – *Non-Clausal Group Theory.*
Group theory examples were taken from (Chang and Lee, 1973), e.g. the statement that a group with the identity element e is commutative if $x * x = e$ holds for every element x. All problems

are represented in a non-clausal form. The original axiomatization and a modified one (Tammet, 1996) are used.

SYJ – *Intuitionistic Syntactic.*
Problems were taken from existing benchmark collections (see Subsection 1.2) and have no obvious semantic interpretation. The ft collection (Sahlin et al., 1992) contains various kind of first-order problems. For some problems the actual abstract or generic problem is added and instances of different sizes included. The JProver collection (Schmitt et al., 2001) contains propositional and first-order problems that are classically valid and can be represented in a pure first-order logic. Type information was removed from all problems. Finally the generic propositional problems from Dyckhoff's benchmark collection are included. For each generic problem the first 20 instances are included in the propositional part and three instances are included in the first-order part of the library.

The syntax of all three added domains was standardized and adapted to the TPTP syntax format. Each problem file was given a header with useful information (see Subsection 2.3). Statistics about these problems can be found in Table III and IV.

The set of problems was split into a first-order part and a propositional part. The first-order part of the ILTP library contains 70 of the propositional problems, but does not contain 204 instances from Dyckhoff's problem collection. This keeps the strong emphasis onto first-order problems.

2.2. Intuitionistic Problem Status and Difficulty Rating

In the TPTP library the difficulty of each problem is rated according to the performance of current (classical) state-of-the-art ATP systems. It expresses the ratio of state-of-the-art systems which are *not* able to solve a problem within a given time limit. For example a rating of 0.0 indicates that every state-of-the-art prover could 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 could not be solved by any state-of-the-art ATP system. This notion of difficulty rating is adapted to the ILTP library. To this end a set of intuitionistic state-of-the-art ATP systems needs to be specified. Comprehensive tests of 14 currently available ATP systems for intuitionistic first-order and propositional logic were performed. For this purpose a test environment was developed for automatically conducting tests and collecting and evaluating the vast amount of data. An eight-processor cluster system was used to simultaneously test several ATP systems on all 2754 problems in the ILTP library. The time limit allowed to solve a problem was set to 600 seconds. The results are published within the ILTP library (see also (Raths, 2005) for the results of release v1.0). The following ATP systems were selected that solve the highest number of problems: the first-order systems ft (C-version) (Sahlin et al., 1992), JProver (Schmitt et al., 2001), ileanTAP(Otten, 1997), ileanCoP(Otten, 2005) for the first-order part, and the propositional systems ft (C-version)[3] (Sahlin et al., 1992), LJT (Dyckhoff, 1992), PITP (Avellone et al., 2004) and STRIP (Larchey-Wendling et al., 2001) for the propositional part of the library.

Each problem was assigned an intuitionistic status. This status is either Theorem, Non-Theorem, Unsolved or Open. Problems with Unsolved status have not been solved under the test condition by any state-of-the-art ATP system.[4] A problem has an Open status if its abstract problem has not been solved at all. No theoretical investigations into the intuitionistic validity of the problems in the TPTP library were done. Instead the intuitionistic status of a problem is marked as Theorem or Non-Theorem if any intuitionistic ATP system was able to prove or refute the problem, respectively. All other problems taken from the TPTP library were given the status Unsolved or Open as described above. Note that problems in the TPTP library whose *classical* status is Satisfiable or Unsatisfiable were negated by the tptp2X tool using the provided format files (see Subsection 2.4).[5]

Table III and Table IV show statistics about the intuitionistic status and rating of the problems in the first-order part and the propositional part of the ILTP library, respectively. Note that the first-order part also contains some propositional problems (see Subsection 2.1).

2.3. Problem Versions, Naming and Presentation

Similar to the TPTP library (Sutcliffe and Suttner, 1998) problems can have alternative presentations resulting in different versions of the original *abstract* problem. There are two main reasons for different versions of a given problem: different axiomatizations and problems that are instances of the same generic (abstract) problem.

The naming of problems was adapted from the TPTP library. Problems are given an unambiguous file name of the form

`DDD.NNN+V[.SSS].p`

consisting of the three letter mnemonic `DDD` of its domain (see Subsection 2.1), the number `NNN` of the abstract problem, the version number

[3] The ft system uses a specialized proof calculus for propositional logic.
[4] In the TPTP library the term Unknown is used instead of Unsolved.
[5] This is the main change in the ILTP release v1.1.1 compared to release v1.1.0.

V, an optional parameter SSS indicating the size of the instance, and an additional letter p. For example SYJ205+1.007.p is the seventh instance of problem number 205 in the domain SYJ.

All files are presented in Prolog syntax and can easily be modified using Prolog. The tptp2X tool (see Subsection 2.4) can be used to convert the syntax to the input syntax of a specific ATP system. Problem files are split into input clauses and have the syntax of Prolog facts. They are marked with a name and a type, which is either axiom, hypothesis, lemma or conjecture. Each problem file includes a header containing information like file name, problem description, status and difficulty rating. This header is marked as a Prolog comment and should not be used by ATP systems to solve the problem. An example file of a problem is given in Figure 1.

```
%--------------------------------------------------------------------------
% File    : GEJ003+1 : ILTP v1.1.1
% Domain  : Constructive Geometry
% Problem : Theorem 3.3. Uniqueness of constructed points
% Version : [P95] axioms.
% English : If two lines are convergent and there is a point that is
%         : incident with both lines, then this point is equivalent to the
%         : intersection point of these lines.

% Refs    : [P95] J. von Plato. The Axioms of Constructive Geometry. Annals
%           of Pure and Applied Logic 76 (2): 169-200, 1995.
% Source  : [P95]
% Names   :

% Status (intuit.) : Theorem
% Rating (intuit.) : 0.75 v1.1.1
% Comments :
%--------------------------------------------------------------------------
include('Axioms/GEJ001+1.ax').
include('Axioms/GEJ002+1.ax').
include('Axioms/GEJ003+1.ax').
include('Axioms/GEJ004+1.ax').
%----------------------------
fof(con_name,conjecture,(
( ! [X,Y,Z] : ((con(X,Y) & ( ~apt(Z,X) & ~apt(Z,Y))) => ~dipt(Z,pt(X,Y))))
)).
%--------------------------------------------------------------------------
```

Figure 1. Example of a problem file (GEJ003+1.p)

More specifically the header files contains the following fields: File contains the problem name and the ILTP release number; Domain identifies the problem domain; Problem and English provide a short and an extended problem description; Version describes the specific version; Refs, Source, Names provide information about references to the problem, the original source and established names of the problem, respectively. The Status (intuit.) and Rating (intuit.) fields specify the *intuitionistic* status and difficulty rating of the problem (see Subsection 2.2), Comment provides additional remarks. The header of problem files taken from the TPTP library additionally includes the fields `Status`, `Rating` and `Syntax`, which specify the *classical* status and difficulty rating, and various syntactic measures of the problem.

2.4. Tools and Prover Database

The TPTP library provides the tptp2X tool for transforming and converting TPTP problem files. This tool can be used for all problems in the ILTP library as well. The tptp2X installation files were adapted to work with the ILTP directory structure. So-called format files were included for all tested intuitionistic ATP systems. They are used together with the tptp2X tool to convert the problems in the ILTP library into the input syntax of the tested ATP systems. The prover database of the library provides information about published intuitionistic ATP systems. For each system some basic information is provided, like author, homepage, short description, references, and test runs on two example problems. A summary and a detailed list of the performance results on running the system on all problems in the ILTP library are given as well.

3. Conclusion

Like the TPTP library for classical logic, the main purpose 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 ATP systems. 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 and propositional logic, which — compared to classical logic — is still in its infancy. The advancement in the field of intuitionistic ATP systems is indicated by the intuitionistic rating of problems, which should gradually decrease as progress is made. Future work includes adding more problems which occur during the practical use of interactive proof assistants like NuPRL (Allen et al., 2000) or Coq (Bertot and Castéran, 2004). Extending the library to other non-classical logics like first-order modal logics or fragments of linear logic is under consideration as well.

Like most problem libraries the ILTP library is an ongoing project. We encourage not only its use but also submissions of new problems suitable for evaluating intuitionistic ATP systems. Users are also invited to submit performance data of their ATP system as well as information about their ATP system itself.

Acknowledgements

The authors would like to thank the anonymous referees for their useful comments.

References

Allen, S., R. Constable, R. Eaton, C. Kreitz, and L. Lorigo. The NuPRL Open Logical Environment. 17^{th} CADE, LNAI 1831, pp. 170–176, Springer, 2000.

Avellone, A., G. Fiorino, and U. Moscato. A new $o(nlogn)$-space decision procedure for propositional intuitionistic logic. *Kurt Gödel Society Collegium Logicum*, volume VIII, pp. 17–33, KGS, 2004.

Balsiger, P., A. Heuerding, and S. Schwendimann. Logics workbench 1.0. *7^{th} TABLEAUX Conference*, LNCS 1397, pp. 35–37, Springer, 1998.

Bertot, Y. and P. Castéran. *Interactive theorem proving and program development*. Texts in Theoretical Computer Science, Springer, 2004.

Chang, C.-L. and R. Lee. *Symbolic logic and mechanical theorem proving*. Academic Press, 1973.

Dafa, L. Replacing the axioms for connecting lines and intersection points by two single axioms. *AAR Newsletter*. 37:1–7, 1997.

Dafa, L. A shorter and intuitive axiom to replace the third axiom of compatibility of equality with apartness and convergence. *AAR Newsletter*. 39:1–6, 1998.

Dyckhoff, R. Contraction-free sequent calculi for intuitionistic logic. *Journal of Symbolic Logic*, 57:795–807, 1992.

Hoos, H. and T. Stützle. SATLIB: An online resource for research on SAT. *SAT 2000*, pp. 283–292, IOS Press, 2000. (`http://www.satlib.org`)

Larchey-Wendling, D., D. Méry, and D. Galmiche. STRIP: Structural sharing for efficient proof-search. *IJCAR-2001*, LNAI 2083, pp. 696–700, Springer, 2001.

Otten, J. ileanTAP: An intuitionistic theorem prover. *6^{th} TABLEAUX Conference*, LNAI 1227, pp. 307–312, Springer, 1997.

Otten, J. Clausal connection-based theorem proving in intuitionistic first-order logic. *TABLEAUX 2005*, LNAI 3702, pp. 245–261, Springer, 2005. (`http://www.leancop.de`)

von Plato, J. The axioms of constructive geometry. *Annals of Pure and Applied Logic*, 76(2): 169–200, 1995.

Raths, T. Evaluating intuitionistic automated theorem provers. Technical Report, University of Potsdam, 2005. (`http://www.iltp.de`)

Raths, T., J. Otten, and C. Kreitz. The ILTP Library: Benchmarking Automated Theorem Provers for Intuitionistic Logic. *TABLEAUX 2005*, LNAI 3702, pp. 333–337, Springer, 2005.

Sahlin, D., T. Franzen, and S. Haridi. An intuitionistic predicate logic theorem prover. *Journal of Logic and Computation*, 2:619–656, 1992.

Schmitt, S., L. Lorigo, C. Kreitz, and A. Nogin. JProver: Integrating connection-based theorem proving into interactive proof assistants. *IJCAR-2001*, LNAI 2083, pp. 421–426, Springer, 2001.

Sutcliffe, G. and C. Suttner. The TPTP problem library - CNF release v1.2.1. *Journal of Automated Reasoning*, 21: 177–203, 1998. (`http://www.tptp.org`)

Tammet, T. A resolution theorem prover for intuitionistic logic. *13^{th} CADE*, LNAI 1104, pp. 2–16, Springer, 1996.

Table III. First-order part of ILTP library v1.1.1: status and rating of problems

Domain	Intuitionistic status: Theorem	Non-Theorem	Open/ Unsolved	Intuitionistic rating: 0.00	0.01 –0.33	0.34 –0.66	0.67 –0.99	1.00
AGT	13	0	39	0	0	2	11	39
ALG	9	0	190	0	3	2	4	190
COM	3	0	0	0	0	0	3	0
CSR	1	0	28	0	0	0	1	28
GEO	6	0	71	0	0	0	6	71
GRA	3	0	15	0	0	0	3	15
GRP	1	2	2	0	0	0	3	2
HAL	0	0	9	0	0	0	0	9
KRS	52	0	106	11	14	7	20	106
LCL	1	2	1	0	1	2	0	1
MGT	25	0	53	5	2	0	18	53
MSC	0	0	2	0	0	0	0	2
NLP	11	0	247	3	4	0	4	247
NUM	24	0	60	0	0	1	23	60
PLA	0	1	5	0	0	0	1	5
PUZ	4	0	2	2	0	0	2	2
SET	76	0	248	12	1	1	62	248
SWC	1	0	422	0	1	0	0	422
SWV	1	5	214	1	0	0	5	214
SYN	120	69	177	101	9	10	69	177
TOP	2	0	1	2	0	0	0	1
GEJ	71	0	22	4	4	10	53	22
GPJ	3	0	2	0	1	0	2	2
SYJ	108	17	3	62	21	19	23	3
$\sum$	535	96	1919	203	61	54	313	1919

Table IV. Propositional part of ILTP library v1.1.1: status and rating of problems

Domain	Intuitionistic status: Theorem	Non-Theorem	Open/ Unsolved	Intuitionistic rating: 0.00	0.01 –0.33	0.34 –0.66	0.67 –0.99	1.00
LCL	0	2	0	2	0	0	0	0
SYN	7	13	0	19	0	0	1	0
SYJ	121	94	37	139	23	28	25	37
$\sum$	128	109	37	160	23	28	26	37