# Problem Libraries for Non-Classical Logics

### – Extended Abstract –

Jens Otten and Thomas Raths

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

## 1 Introduction

Problem libraries for automated theorem proving (ATP) systems play a crucial role when developing, testing, benchmarking, and evaluating ATP systems and proof search calculi for classical and non-classical logics. During the development of ATP systems problem libraries are necessary for *testing*, providing evidence about the correctness of implemented proof search calculi. Furthermore, new proof search techniques and strategies can be evaluated, improving the efficiency of the implemented algorithms. Theoretical investigations into the time complexity of proof search calculi and their implementations are difficult and usually not carried out. Hence, standardized problem libraries are necessary to measure their performance and put the *evaluation* of ATP systems on a firm basis. The performance evaluation of ATP systems provides evidence about the efficiency of the different underlying proof search calculi, which is important in order to measure and to make progress in the field of ATP. Finally, problem libraries provide a platform for users of ATP systems to submit problems that occur in their applications. This, in turn, allows developers to use these problems to further improve the efficiency of their ATP systems.

For *classical* first-order logic, the first (printed) problem collections started 1976 with 63 problems by McCharen, Overbeek and Wos [6] and was ten years later extended by Pelletier's "Seventy-five Problems for Testing Automatic Theorem Provers" [9]. In 1997, version 2.0.0 of the TPTP library for classical logic included the first 217 problems in first-order (i.e. non-clausal) form [14]. Since then its number has risen significantly and the most recent version 6.0.0 of the TPTP library contains almost 8000 first-order problems (in non-clausal form) divided into 36 different problem domains. The existence of the TPTP library has stimulated the development of efficient ATP systems for classical logic leading to significant performance improvements.

For *non-classical* logics, such as intuitionistic and modal logic, only few problem libraries were developed so far. One of the first ATP systems for first-order *intuitionistic* logic was tested on a small collection of 43 first-order problems [12]. As intuitionistic logic has the same syntax as classical logic, problems used to test classical ATP systems can basically also be used for testing intuitionistic ATP systems. For *modal* logic the situation is more complicated, as it extends classical logic by the unary modal operators $\Box$ and $\Diamond$. The approach of generating random problem formulae for *propositional* modal logic is not practical for *first-order* modal logic. A small collection of problems for first-order modal logic was developed for testing one of the first ATP systems for first-order modal logic [15].

The main purpose of this paper is to make people aware of existing problem libraries for non-classical logic and to discuss ongoing and future work on these libraries. The paper provides an overview of existing problem libraries for some important non-classical logics, namely for first-order intuitionistic and first-order modal logics, and describes possible extensions and ideas for future developments of problem libraries for non-classical logics.

# 2 Problem Libraries for Intuitionistic and Modal Logics

The ILTP library and the QMLTP library provide a platform for testing and evaluating ATP systems for first-order intuitionistic and first-order modal logics, respectively.

## 2.1 The ILTP Library for Intuitionistic Logic

The current release v1.1.2 of ILTP library includes a total number of 2754 problems, which are grouped into 24 problem domains [11]. It is available online at `http://www.iltp.de`. Due to the ongoing interest in *propositional* intuitionistic logic, the problem set was split into a first-order part, which contains 2550 problems (2480 first-order problems and 70 propositional problems), and a propositional part, which contains 274 (propositional) problems.

**Contents.** 2324 problems were taken from the set of "FOF" problems in non-clausal form of the TPTP library for first-order classical logic [14]. These problems are classified into 21 domains, named after their corresponding name in the TPTP library. Among these problem domains are, e.g., AGT (agents), ALG (general algebra), GEO (geometry), KRS (knowledge representation), MGT (management), NLP (natural language processing), NUM (number theory), SET (set theory), SWC (software creation), SWV (software verification), and SYN (syntactic). Additionally, the three domain GEJ, GPJ, and SYJ with a total number of 430 problems were created. These domains contain problems from constructive geometry, group theory, and syntactic intuitionistic problems. In general, each problem file consists of a set of axioms $Ax_1,\ldots,Ax_n$ and a conjecture *Conj*. Out of the 2754 problems, 1106 problems contain up to 9 axioms, 1485 problems contain between 10 and 99 axioms, and 163 problems contain between 100 and 922 axioms.

**Syntax and Representation.** The *syntax* for representing formulae and the naming scheme for the problem files were adopted from the TPTP library [14]. Each problem file is given an unambiguous name of the form `DDD.NNN+V[.SSS].p` consisting of its domain `DDD`, the number `NNN` of the problem, its version number `V`, and an optional parameter `SSS` indicating the size of the instance. Each problem file includes a header with useful information, e.g., the file name, a problem description, references, sources, and the intuitionistic status and rating. The ILTP library contains several *format files* that can be used to translate the problems in the ILTP library into the input syntax of existing ATP systems for intuitionistic logic.

**Intuitionistic Status and Rating.** The *intuitionistic status* of a problem formula is either `Theorem`, `Non-Theorem`, `Unsolved`, or `Open`. If at least one intuitionistic ATP system has proved or refuted the problem formula $(Ax_1 \wedge \ldots \wedge Ax_n) \Rightarrow Conj$, the status is `Theorem` or `Non-Theorem`, respectively. Problems with an `Unsolved` status have not yet been solved by any intuitionistic ATP system, but it is known if they are valid or invalid; for problems with an `Open` status it is not known if they are valid or invalid. The *intuitionistic rating* of a problem indicates the difficulty of that problem with respect to current (intuitionistic) state-of-the-art ATP systems.

## 2.2 The QMLTP Library for Modal Logics

The current release v1.1 of the QMLTP library includes a total number of 600 problems, which are grouped into 11 problem domains [10]. It is available online at `http://www.iltp.de/qmltp/`. Out of these 600 problems, 580 are unimodal problems and 20 are multimodal problems; 421 problems are first-order problems, 179 are propositional problems.

**Contents.** The 245 problems in the domains GAL, GLC, GNL, GSE, GSV, and GSY were generated by using Gödel's embedding of intuitionistic logic into the modal logic S4 [3] The original problems were taken from the domains ALG (general algebra), LCL (logic calculi), NLP (natural language processing), SET (set theory), SWV (software verification), and SYN (syntactic) of the TPTP library [14]. The domains APM (applications mixed) and SYM (syntactic modal) contain 255 problems from different applications (planning, querying databases, natural language processing and software verification), syntactic problems from various textbooks, and problems from the TANCS-2000 system competition for modal (propositional) ATP systems [5]. The domains NLP (natural language processing) and SET (set theory) contain a set of 80 classical first-order problems, i.e. problems containing no modal operators, from the corresponding domains of the TPTP library [14]. Furthermore, the domain MML (multimodal logic) contains 20 problems in a multimodal logic syntax from various textbooks and applications. Again, each problem file consists of a set of axioms $Ax_1,\ldots,Ax_n$ and a conjecture *Conj*. Out of the 600 problems, 342 problems contain up to 9 axioms, 210 problems contain between 10 and 99 axioms, and 48 problems contain between 100 and 150 axioms.

**Syntax and Representation** The naming scheme of the problem files is adopted from the TPTP library. For the *syntax* of the modal problems, the syntax of the TPTP library [14] was extended by the two operators `#box:` and `#dia:` representing the modal operators $\Box$ and $\Diamond$, respectively. For multimodal logic, the modal operators $\Box_i$ and $\Diamond_i$ are represented by `#box(i):` and `#dia(i):`, respectively. In this case the command `set_logic` of the TPTP *process instruction language* is used to specify the semantics of the modal operators. E.g., `tpi(1,set_logic,modal([cumulative], [(1,d),(2,s4)]))` specifies that the multimodal operators $\Box_1/\Diamond_1$ and $\Box_2/\Diamond_2$ are interpreted with respect to the cumulative domain variant of the modal logics D and S4, respectively. Each problem file includes a header with useful information, e.g., the file name, a problem description, references, sources, and the modal status and rating. *Format files* are included in the QMLTP library that can be used to convert the problems in the QMLTP library into the input syntax of existing ATP systems for (first-order) modal logic.

**Modal Status and Rating.** The *modal status* of a problem is either `Theorem`, `Non-Theorem`, or `Unsolved` and is specified with respect to a particular modal logic and *domain condition*. Release v1.1 of the QMLTP library provides modal status and rating information for the modal logics K, D, T, S4, and S5 with constant, cumulative, and varying domain conditions. Term designation is assumed to be rigid, i.e., terms denote the same object in each world, and terms are local, i.e., any ground term denotes an existing object in every world. Even though the modal status and rating information is only provided for the standard modal logics K, D, T, S4, and S5, the usage of the problem library is not restricted to these logics. Furthermore, the *local* logical consequence is used (see also Section 3.1): the status of a problem is `Theorem` if the (modal) formula $(Ax_1 \wedge \ldots \wedge Ax_n) \Rightarrow Conj$ was proven valid by at least one (modal) ATP system; if the formula was refuted, i.e. proven invalid, its status is `Non-Theorem`.

# 3 Ongoing Work and Future Plans

Current and future work include extending the existing problem libraries and developing a common benchmark platform for non-classical logics.

## 3.1 Extending the ILTP Library and the QMLTP Library

The ongoing work on the ILTP library and the QMLTP library includes increasing the number of problems and providing status and rating information for additional modal logics/semantics.

**Adding More Difficult Problems.** In order to make meaningful performance evaluations possible, problem libraries need to contain a sufficient amount of challenging, i.e., difficult problems.

Out of the 2550 problems of the first-order part of ILTP library v1.1.2, approximately 1000 problems (40%) are solved, i.e., proved or refuted by current state-of-the-art ATP systems for intuitionistic logic. Hence, the problems in the ILTP library are still sufficiently difficult in order to support the development of more efficient ATP systems. Out of the 274 problems of the propositional part, around 95% are solved by current propositional state-of-the-art ATP systems [4]. Even though the focus of the ILTP library is on first-order problems, more propositional problems will be added in the future, e.g., by generating non-clausal propositional instances of first-order formulae using the first2p tool [1].

Out of the 580 unimodal problems of the QMLTP library v1.1, between 60% (modal K) and 85% (modal S5) are solved by current state-of-the-art ATP systems for modal logic. Additional (syntactic) problems are obtained by applying Gödel's embedding for intuitionistic logic to more difficult problems.

**Semantics of Modal Operators and Logical Consequence.** The information about the modal status and rating of the problems in the QMLTP library will be extended to cover additional logics, i.e., the modal logics K4, D4, and B. Currently, the only ATP systems that support these logics use an embedding into higher-order logic [1]. Even if the *information* about modal status and rating for these logics is currently not included in the QMLTP library, users can already test and benchmark their ATP systems for these logics by using the existing problem sets in the library.

Currently, all modal status and rating information are based on the *local* logical consequence of modal logic. Let $M \models_w F$ iff $M$ is a (Kripke-)model for the formula $F$ in the world $w$ and let $M \models_w \Gamma$ iff $M \models_w F$ for all $F \in \Gamma$. The *local logical consequence* $\models^L$ is defined as $\Gamma \models^L F$ iff for every model $M$ and every world $w$: if $M \models_w \Gamma$ then $M \models_w F$ holds. The formulae (i.e. Axioms) in the set $\Gamma$ are called *local hypotheses* and this kind of notion is often used in the literature. In this case the *deduction theorem* holds, i.e. $\Gamma \cup F' \models^L F$ iff $\Gamma \models^L (F' \Rightarrow F)$ (see also the remarks in the last paragraph of Section 2.2). The *global logical consequence* $\models^G$ is defined as $\Delta \models^G F$ iff for every model $M$: if $M \models_w \Delta$ for every world $w$ then $M \models_w F$ for every world $w$ holds; the formulae in the set $\Delta$ are called *global hypotheses*. Modal status and rating information using the *global* logical consequence will be included in future versions. Until then, users can already test their ATP systems using the global logical consequence (see, e.g.,[2]), but are expected to indicate whenever they diverge from the standard local logical consequence.

**Problems from Applications.** Future versions of the problem libraries will include more problems from applications. All users who have an application that uses (first-order) intuitionistic or modal logic are asked to get in contact with the authors and submit their problems to the ILTP or QMLTP library.

## 3.2 A Common Benchmark Platform for Non-Classical Logics

In order to better support the development of ATP systems for non-classical logics, the development of a common benchmark platform for non-classical logics, similar to the StarExec platform, is currently discussed. StarExec [13] provides an infrastructure to researchers to manage benchmark libraries and solvers/ATP systems for some popular logics, such as first-order classical logic and propositional classical logic (SAT). A common benchmark platform for non-classical logics would not only cover intuitionistic and modal logics, but also less popular non-classical logics, such as interval or alternating-time temporal logics, for which the development of separate, independent platforms is not practical (and which are still not widespread enough to include them in the StarExec platform). It would provide a forum for collecting benchmark problems that are currently independently developed, collected and used by developers of non-classical ATP systems, and would make it easier for researchers developing applications to submit their non-classical problems.

## 4 Conclusion

Non-classical logics, such as intuitionistic and modal logics, have many applications, e.g., in Computer Science, Artificial Intelligence and Philosophy. In contrast to classical logic, the development of ATP systems for non-classical is still in its infancy. This is in particular true for most *first-order* non-classical logics. Problem libraries play an important role when developing, testing, and evaluating ATP systems. They stimulate the development of novel, more efficient calculi and ATP systems. The ILTP library and the QMLTP library are comprehensive and established problem libraries for first-order intuitionistic and first-order modal logics. They already supported the development of some of the fastest ATP systems for first-order intuitionistic and several first-order modal logics [1, 7, 8]. Future versions of these libraries will include more problems as well as status and rating information for further (modal) logics. A common benchmark platform for non-classical logics will support additional non-classical logics.

**Acknowledgements.** The authors would like to thank Christoph Benzmüller and Geoff Sutcliffe for many helpful discussions and suggestions.

## References

[1] C. Benzmüller, J. Otten, T. Raths. Implementing and Evaluating Provers for First-order Modal Logics. In L. De Raedt et al., Eds., *20th European Conference on Artificial Intelligence, ECAI 2012*, pp. 163–168. IOS Press, Amsterdam, 2012.

[2] C. Benzmüller, T. Raths. HOL Based First-Order Modal Logic Provers. In K. McMillian et al., Eds., *LPAR 2013*, LNCS 8312, pp. 127–136. Springer, Heidelberg, 2013.

[3] K. Gödel. An Interpretation of the Intuitionistic Sentential Logic. In J. Hintikka, Ed., *The Philosophy of Mathematics*, pp. 128–129. Oxford University Press, Oxford, 1969.

[4] R. Goré, J. Thomson. An Improved BDD Method for Intuitionistic Propositional Logic: BDDIntKt System Description. In M. P. Bonacina, Ed., *CADE 2013*, LNCS 7898, pp. 275–281. Springer, Heidelberg, 2013.

[5] F. Massacci, F.M. Donini: Design and Results of TANCS-2000 Non-classical (Modal) Systems Comparison. In R. Dyckhoff., Ed., *TABLEAUX 2000*, LNAI 1847, pp. 50–56. Springer, Heidelberg, 2000.

[6] J.D. McCharen, R.A. Overbeek, L.A. Wos. Problems and Experiments for and with Automated Theorem-Proving Programs. *IEEE Transactions on Computers*, C-25(8):773–782, 1976.

[7] J. Otten. leanCoP 2.0 and ileanCoP 1.2: High Performance Lean Theorem Proving in Classical and Intuitionistic Logic. In A. Armando et al., Eds., *IJCAR 2008*, LNCS 5195, pp. 283–291. Springer, Heidelberg, 2008.

[8] J. Otten. MleanCoP: A Connection Prover for First-Order Modal Logic. In S. Demri, D. Kapur, C. Weidenbach, Eds., *IJCAR 2014*, LNAI 8562, pp. 269–276. Springer, Heidelberg, 2014.

[9] F.J. Pelletier. Seventy-five Problems for Testing Automatic Theorem Provers. *Journal of Automated Reasoning* 2(2):191–216, 1986.

[10] T. Raths, J. Otten. The QMLTP Problem Library for First-order Modal Logics. In B. Gramlich, D. Miller, U. Sattler, Eds., *IJCAR 2012*, LNAI 7364, pp. 454–461. Springer, Heidelberg, 2012.

[11] T. Raths, J. Otten, C. Kreitz. The ILTP Problem Library for Intuitionistic Logic. *Journal of Automated Reasoning*, 38(1–3): 261–271, 2007.

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

[13] A. Stump, G. Sutcliffe, C. Tinelli. Introducing StarExec: a Cross-Community Infrastructure for Logic Solving. In V. Klebanov et al., Eds., *COMPARE 2012*, Workshop Proceedings, p. 2, Manchester/UK, 2012.

[14] G. Sutcliffe. The TPTP Problem Library and Associated Infrastructure: The FOF and CNF Parts, v3.5.0. *Journal of Automated Reasoning*, 43(4):337–362, 2009.

[15] V. Thion, S. Cerrito, M. Cialdea Mayer. A General Theorem Prover for Quantified Modal Logics. In U. Egly, C.G. Fermüller, Eds., *TABLEAUX-2002*, LNCS 2381, pp. 266–280. Springer, Heidelberg, 2002.