The QMLTP Library: Benchmarking Theorem Provers for Modal Logics

Thomas Raths[⋆] Jens Otten

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

Abstract. The Quantified Modal Logic Theorem Proving (QMLTP) library provides a platform for testing and evaluating automated theorem proving (ATP) systems for first-order modal logics. The main purpose of the library is to foster the development of new ATP systems and to put their comparison onto a firm basis. The current version 1.0.1 of the QMLTP library includes 500 problems represented in a standardized extended TPTP syntax. Status and difficulty rating for all problems were determined by running comprehensive tests with currently available ATP systems. In the current version of the library the modal logics D and S4 with constant and cumulative domains are considered.

1 Introduction

Testing *automated theorem proving* (ATP) systems using standardized problem sets is a well-established method for measuring their performance. Popular examples are the TPTP library [25] for classical logic and the ILTP library [20] for intuitionistic logic. These libraries have stimulated the development of more efficient ATP systems for these logics. The aim of the *Quantified Modal Logic Theorem Proving (QMLTP) library* is to provide a comprehensive set of problems for various first-order modal logics. This will put the testing and evaluation of ATP systems for first-order modal logic on a firm basis, make meaningful system evaluations and comparisons possible, and will allow to measure practical progress in the field.

There already exist a few benchmark problems and methods for some *propositional* modal logics, e.g. there are some scalable problem classes [2] and procedures that generate formulas randomly in a normal form [18]. For *first-order* modal logics, there are only small collections of formulas available, e.g. formulas used for testing the ATP system GQML-Prover [27]. Version 5.1.0 of the TPTP library contains 187 modal syntactic problems as well, mostly from textbooks and formulated in a typed higher-order language. All modal first-order problems are included in the QMLTP library as well.

Modal logics extend classical logic with the modalities "it is necessarily true that" and "it is possibly true that" represented by the unary operators $\Box$ and $\Diamond$, respectively. The (Kripke) semantics of modal logics is defined by a set of worlds constituting classical logic interpretations, and a binary accessibility relation on this set. Then $\Box F$ or

[⋆] This work is partly funded by the German Science Foundation DFG under reference number KR858/9-1.

$\Diamond F$ is true in a world w, if F is true in *all* worlds accessible from w or *some* world accessible from w, respectively. First-order or *quantified* modal logics (QMLs) extend propositional modal logics by *domains* specifying sets of objects that are associated with each world, and the standard universal and existential quantifiers [5, 12].

First-order modal logics allow a natural and compact knowledge representation. The subtle combination of the modal operators and first-order logic enables specifications on epistemic, dynamic and temporal aspects, and on infinite sets of objects. For this reason, first-order modal logics have many applications, e.g. in planning, natural language processing, program verification, querying knowledge bases, and modeling communication. In these applications, modalities are used to represent incomplete knowledge, programs, or to contrast different sources of information. All these applications would benefit from a higher degree of automation. Consequently there is a real need for efficient ATP systems for first-order modal logic.

This paper introduces release v1.0 of the QMLTP library. It describes how to obtain and use the QMLTP library, provides details about the contents of the library, and information about status, difficulty rating and syntax of the included problem sets.

1.1 Obtaining and Using the Library

The QMLTP library is available at `http://www.iltp.de/qmltp`. It is structured into four subdirectories:

`Axioms`	- contains the axiom files.
`Documents`	- contains papers, statistic files, and other documents.
`Problems`	- contains a directory for each domain with problem files.
`TPTP2X`	- contains the tptp2X tool and the format files.

There are a few important conditions that should be observed when presenting results of ATP systems based on the QMLTP library. The release number of the QMLTP library and the version of the tested ATP system including all settings must be documented. Each problem should be referred to by its unique name and 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 Section 2.4). The header information of each problem may not be exploited by an ATP system.

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 QMLTP library are given. This makes the verification and validation of the given performance data possible.

2 Contents of the QMLTP Library

Figure 1 provides a summary of the contents of release v1.0 of the QMLTP library.

Number of problem domains	8	
Number of problems	500	(100%)
Number of first-order problems	325	(65%)
Number of propositional problems	175	(33%)
Number of problems with equality	163	(33%)
Number of pure equality problems	16	(3%)

Table 1. Overall statistics of the QMLTP library v1.0

2.1 The QMLTP Domain Structure

The 500 problems of the QMLTP library are divided into eight problem domains. These domains are APM, GAL, GLC, GNL, GSE, GSV, GSY, and SYM.

1. APM – *applications mixed*.
   10 problems from planning, querying databases, natural language processing and communication, and software verification [6–8, 21, 23, 24].[1]
2. GAL/GLC/GNL/GSE/GSV/GSY – *Gödel's embedding*.
   245 problems are generated by using Gödel's embedding of intuitionistic logic into the modal logic S4 [14]. The original problems were taken from the TPTP library [25] and derived from problems in the domains ALG (general algebra), LCL (logic calculi), NLP (natural language processing), SET (set theory), SWV (software verification), and SYN (syntactic), respectively.
3. SYM – *syntactic modal*.
   175 problems from various textbooks [9–11, 13, 19, 22, 27] and 70 problems from the TANCS-2000 system competition for modal ATP systems [16].

2.2 Modal Problem Status and Difficulty Rating

As already done in the TPTP and ILTP library, each problem is assigned a status and a rating. The *rating* determines the difficulty of a problem with respect to current state-of-the-art ATP systems. It is the fraction of state-of-the-art ATP systems which are *not* able to solve a problem within a given time limit. For example a rating of 0.3 indicates that 30% of the state-of-the-art systems do *not* solve the problem; a problem with rating of 1.0 cannot be solved by any state-of-the-art system. A *state-of-the-art* system is an ATP system whose set of solved problems is not subsumed by any another ATP system.

Each problem is assigned a modal *status*. It is either `Theorem`, `Non-Theorem` or `Unsolved`. For problems with `Theorem` status at least one of the considered ATP systems has found a proof or counter model (refutation), respectively. Problems with `Unsolved` status have not been solved by any ATP system. The status is specified with respect to a particular modal logic, e.g. D, K, K4, D4, S4, T, and a particular domain (condition), i.e. constant, cumulative, or varying domains.

[1] There are only few problems from real applications in the current release of the QMLTP library. Future versions will include more problems from applications once they get submitted to the QMLTP library.

		Modal Status			Modal Rating				
Logic	**Domain**	Theorem	Non-Theorem	Unsolved	0.00	0.01–0.49	0.50–0.99	1.00	$\sum$
D	cumulative	135	255	110	46	56	288	110	500
	constant	151	242	107	44	54	295	107	500
S4	cumulative	274	116	110	81	89	219	111	500
	constant	300	107	93	78	84	231	107	500

Table 2. QMLTP library v1.0.1: status and rating summary

When determining the modal status, the standard semantics of first-order modal logics [10, 12] is considered. Note that the problem files of the QMLTP library are primary intended to present the *syntax* of modal formulas. The options of the intended *semantics*, e.g. the interpretation of the modal operators in the different modal logics, is left to the (user of the) particular ATP system. For the first release v1.0 of the QMLTP library all rating and status information is with respect to the first-order modal logics D and S4 with constant and cumulative domains. Term designation is assumed to be rigid, i.e. terms denote the same object in each world, and terms shall be local, i.e. any ground term denotes an existing object for each world. Future versions of the library will consider further modal logics as well.

To determine rating and status of the problems, the following ATP systems were used.[2] For constant domains: Leo-II 1.2, Satallax 1.4, f2p+MSPASS 3.0, MleanTAP 1.2, MleanSeP 1.2 and MleanCoP 1.2; for cumulative domains: f2p+MSPASS 3.0, MleanTAP 1.2, MleanSeP 1.2 and MleanCoP 1.2.[3] Leo-II [4] and Satallax [1] are ATP systems for typed higher-order logic.[4] Leo-II uses an extensional higher-order resolution calculus. Satallax uses a complete ground tableau calculus. Both ATP systems use an embedding of quantified modal logic into simple type theory [3]. MleanTAP and MleanSeP are compact ATP systems for several first-order modal logics. MleanTAP implements an analytic free-variable tableau using prefix unification similar to the ileanTAP system [17] for intuitionistic logic. MleanSeP implements a modular analytic sequent calculus. MleanCoP uses the clausal connection calculus and extends the classical leanCoP prover by adding prefixes to the literals and a prefix unification algorithm. The system f2p+MSPASS implements an instance-based method to generate ground formulas that are proved by MSPASS [15] in propositional modal logic

Table 2 shows statistics about the modal status and rating of the problems in the current version of the QMLTP library for all modal logics under consideration. For the modal logic S4 with cumulative domains Table 3 provides this information for each domain in the library.

[2] All systems were run on a 3.4 GHz Xeon system using Linux and ECLiPSe Prolog 5.10. The time limit for all proof attempts was set to 600 seconds.

[3] These are the only correct ATP systems currently available for first-order modal logic.

[4] These two higher-order ATP system were selected as they solve the highest number of problems in the last CASC-J5 [26].

	Modal Status			Modal Rating					
Domain	Theorem	Non-Theorem	Unsolved	0.00	0.25	0.50	0.75	1.00	$\sum$
APM	6	3	1	1	3	3	2	1	10
GAL	1	0	9	0	1	0	0	9	10
GLC	8	1	16	0	0	1	8	16	25
GNL	5	0	5	0	1	4	0	5	10
GSE	56	0	19	0	5	6	45	19	75
GSV	36	1	13	0	19	12	6	13	50
GSY	45	22	8	9	24	11	23	8	75
SYM	117	89	39	71	36	22	76	40	245
$\sum$	274	116	110	81	89	59	160	111	500

Table 3. QMLTP library v1.0.1: status and rating summary for S4 with cumulative domains

2.3 Naming, Syntax and Presentation

Similar to the TPTP library, each problem is given an unambiguous name. The problem *name* has the form `DDD.NNN+V[.SSS].p` consisting of the mnemonic `DDD` of its domain, the number `NNN` of the problem, its version number `V`, and an optional parameter `SSS` indicating the size of the instance. For example `SYM001+1.p` is (the first version of) the first problem in the domain SYM.

For the *syntax* of the problems the Prolog syntax of the TPTP library [25] is extended by the modal operators. We use the two Prolog atoms "`#box`" and "`#dia`" for representing $\Box$ and $\Diamond$, respectively. The formulas $\Box F$ and $\Diamond F$ are then represented by "`#box:F`" and "`#dia:F`", respectively (see also Figure 1). For future extensions to multi-modal logic these atoms can, e.g., be used in Prolog terms of the form "`#box(i)`" or "`#dia(i)`" in which the index "`i`" is an arbitrary Prolog atom. As there exists no ATP system for first-order multi-modal logic, the current release of the QMLTP-Library is restricted to uni-modal problems only.

A header with useful information is added to the *presentation* of each problem. It is adapted from the TPTP library and includes information about the file name, the problem description, the modal status and the modal difficulty rating. An example file of a problem is given in Figure 1.

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 QMLTP library as well. *Format files* were included for all tested modal ATP systems. They are used together with the tptp2X tool to convert the problems in the QMLTP library into the input syntax of the tested ATP systems. The prover database of the library provides information about published modal 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 QMLTP library are given as well.

```
%------------------------------------------------------------------------
% File     : SYM001+1 : QMLTP v1.0.1
% Domain   : Syntactic (modal)
% Problem  : Barcan scheme instance. (Ted Sider's qml wwf 1)
% Version  : Especial.
% English  : if for all x necessarily f(x), then it is necessary that for
%            all x f(x)
% Refs     : [Sid09] T. Sider. Logic for Philosophy. Oxford, 2009.
%          : [Brc46] [1] R. C. Barcan. A functional calculus of first
%            order based on strict implication. Journal of Symbolic Logic
%            11:1-16, 1946.
% Source   : [Sid09]
% Names    : instance of the Barcan formula
%
% Status   :      cumulative  constant
%            D    Unsolved    Theorem       v1.0.1
%            S4   Unsolved    Theorem       v1.0.1
%
% Rating   :      cumulative  constant
%            D    1.00         0.00        v1.0.1
%            S4   1.00         0.00        v1.0.1
%
% term conditions for all terms:  designation: rigid,  extension: local
%
% Comments :
%------------------------------------------------------------------------
```

Fig. 1. Example of a problem file (`SYM001+1`)

3 Conclusion

The first official version of a problem library for first-order modal logic was presented. Like the TPTP library for classical logic and the ILTP library for intuitionistic logic, the objective of the QMLTP library is to put the testing and evaluation of ATP systems for first-order modal logic onto a firm basis. It will make meaningful systems evaluations and comparisons possible and help to ensure that published results reflect the actual performance of an ATP system. Experiences with existing libraries have shown that they stimulate the development of novel, more efficient calculi and implementations. Future work includes adding more problems that are used within applications and the extension to, e.g., some first-order multi-modal logics.

Like other problem libraries the QMLTP library is an ongoing project. We invite all interested users to submit problems and ATP systems that use first-order modal logics.

Acknowledgements. The authors would like to thank the following people who gave valuable feedback on a preliminary version of the QMLTP library, contributed problems and ATP systems, and were involved in discussions about the TPTP syntax for modal logic: Geoff Sutcliffe, Christoph Benzmüller, Silvio Ghilardi, Rajeev Goré, John Halleck, Ullrich Hustadt, Renate Schmidt, Stephan Schulz, and Russell Wallace.

References

1. J. Backes, C. E. Brown. Analytic Tableaux for Higher-Order Logic with Choice. *IJCAR 2010*, LNCS 6173, pp. 76–90. Springer, 2010.

2. P. Balsiger, A. Heuerding, S. Schwendimann. A Benchmark Method for the Propositional Modal Logics K, KT, S4. *Journal of Automated Reasoning*, 24:297–317, 2000.
3. C. Benzmüller, L. Paulson. Quantified Multimodal Logics in Simple Type Theory. Seki Report SR-2009-02 (ISSN 1437-4447), Saarland University, 2009.
4. C. Benzmüller, L. Paulson, F. Theiss, A. Fietzke. LEO-II – A Cooperative Automatic Theorem Prover for Higher-Order Logic. *IJCAR 2008*, LNAI 5195, pp. 162–170. Springer, 2008.
5. P. Blackburn, J. van Bentham, F. Wolter. *Handbook of Modal Logic*. Elsevier, 2006.
6. V. Boeva, L. Ekenberg. A Transition Logic for Schemata Conflicts. *Data & Knowledge Engineering*, 51(3):277–294, 2004.
7. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, R. Rosati. EQL-Lite: Effective First-Order Query Processing in Description Logics. In M. M. Veloso, Ed., *IJCAI-2007*, pp. 274–279, 2007.
8. L. Fariñas del Cerro, A. Herzig, D. Longin, O. Rifi. Belief Reconstruction in Cooperative Dialogues. *AIMSA 1998*, LNCS 1480, pp. 254–266. Springer, 1998.
9. M. Fitting. *Types, Tableaus, and Goedel's God*. Kluwer, 2002.
10. M. Fitting, R. L. Mendelsohn. *First-Order Modal Logic*. Kluwer, 1998.
11. G. Forbes. *Modern Logic. A Text in Elementary Symbolic Logic*. Oxford University Press, 1994.
12. J. Garson. Quantification in Modal Logic. *Handbook of Philosophical Logic*, volume II, pp. 249–307. D. Reidel Publ. Co, 1984.
13. R. Girle. *Modal Logics and Philosophy*. Acumen Publ., 2000.
14. K. Gödel. An Interpretation of the Intuitionistic Sentential Logic. In J. Hintikka, Ed., *The Philosophy of Mathematics*, pp. 128–129. Oxford University Press, 1969.
15. U. Hustadt, R. A. Schmidt MSPASS: Modal Reasoning by Translation and First-Order Resolution. R. Dyckhoff., Ed., *TABLEAUX-2000*, LNAI 1847, pp. 67–81. Springer, 2000.
16. F. Massacci, F. M. Donini: Design and Results of TANCS-2000 Non-classical (Modal) Systems Comparison. R. Dyckhoff., Ed., *TABLEAUX-2000*, LNAI 1847, pp. 50–56. Springer, 2000.
17. J. Otten. ileanTAP: An Intuitionistic Theorem Prover. In D. Galmiche, Ed., *TABLEAUX-97*, LNAI 1227, pp. 307–312. Springer, 1997.
18. P. F. Patel-Schneider, R. Sebastiani. A New General Method to Generate Random Modal Formulae for Testing Decision Procedures. *Journal of Articial Intelligence Research*, 18:351–389, 2003.
19. S. Popcorn. *First Steps in Modal Logic*. Cambridge University Press, 1994.
20. T. Raths, J. Otten, C. Kreitz. The ILTP Problem Library for Intuitionistic Logic. *Journal of Automated Reasoning*, 38(1–3): 261–271, 2007.
21. R. Reiter. What Should a Database Know? *Journal of Logic Programming* 14(1–2):127–153, 1992.
22. T. Sider. *Logic for Philosophy*. Oxford University Press, 2009.
23. M. Stone. Abductive Planning With Sensing. In *AAAI-98*, pp. 631–636. Menlo Park CA., 1998.
24. M. Stone. Towards a Computational Account of Knowledge, Action and Inference in Instructions. *Journal of Language and Computation*, 1:231–246, 2000.
25. 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.
26. G. Sutcliffe. The 5th IJCAR automated theorem proving system competition – CASC-J5. *AI Communications*, 24(1):75–89, 2011.
27. 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, 2002.