The QMLTP Problem Library for First-Order Modal Logics

Thomas Raths* and Jens Otten

Institut für Informatik, University of Potsdam
August-Bebel-Str. 89, 14482 Potsdam-Babelsberg, Germany
{traths,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 stimulate the development of new modal ATP systems and to put their comparison onto a firm basis. Version 1.1 of the QMLTP library includes 600 problems represented in a standardized extended TPTP syntax. Status and difficulty rating for all problems were determined by running comprehensive tests with existing modal ATP systems. In the presented version 1.1 of the library the modal logics K, D, T, S4 and S5 with constant, cumulative and varying domains are considered. Furthermore, a small number of problems for multi-modal logic are included as well.

1 Introduction

Problem libraries are essential tools when developing and testing *automated theorem proving* (ATP) systems for various logics. Popular examples are the TPTP library [23] for classical logic and the ILTP library [18] for intuitionistic logic. These libraries help many developers to test and improve their ATP system and, hence, have stimulated the development of more efficient ATP systems.

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. $\Box F$ or $\Diamond F$ are 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 extend propositional modal logics by *domains* specifying sets of objects that are associated with each world, and the standard universal and existential quantifiers [4, 10].

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 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,

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

B. Gramlich, D. Miller, and U. Sattler (Eds.): IJCAR 2012, LNAI 7364, pp. 454–461, 2012.

or to contrast different sources of information. 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 whose development is still in its infancy.

The *Quantified Modal Logic Theorem Proving (QMLTP) library* provides a comprehensive set of standardized problems in first-order modal logic and, thus, constitutes a basis for testing and evaluating the performance of ATP systems for modal logics. The main purpose of the QMLTP library is to stimulate the development of new calculi and ATP systems for first-order modal logics. It puts the testing and evaluation of ATP systems for first-order modal logics on a firm basis, makes meaningful system evaluations and comparisons possible, and allows 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 [1] and procedures that generate formulas randomly in a normal form [16]. For *first-order* modal logics, there are only small collections of formulas available, e.g., a small set of formulas used for testing the ATP system GQML-Prover [25]. Version 5.1.0 of the TPTP library also contains some modal problems, mostly from textbooks, formulated in a typed higher-order language. All these existing sets of problems are included in the QMLTP library.

This paper introduces release v1.1 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 problems.

2 Obtaining and Using the Library

The QMLTP library is available online at `http://www.iltp.de/qmltp/`. The main file containing the modal problems is structured into three subdirectories:

`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 modal ATP systems based on the QMLTP library; see [23]. 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 3.4). The header information (see Section 3.3) of each problem may not be exploited by an ATP system.

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

3 Contents of the QMLTP Library

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

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

Number of problem domains	11	
Number of problems	600	(100%)
Number of first-order problems	421	(70%)
Number of propositional problems	179	(30%)
Number of uni-modal problems	580	(97%)
Number of multi-modal problems	20	(3%)

3.1 The QMLTP Domain Structure

The 600 problems of the library are divided into problem classes or *domains*. These domains are APM, GAL, GLC, GNL, GSE, GSV, GSY, MML, NLP, SET, and SYM.[1]

1. APM – *applications mixed*.
   10 problems from planning, querying databases, natural language processing and communication, and software verification [5, 7, 8, 19, 21, 22].
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 [13]. The original problems are from the following domains of the TPTP library: ALG (general algebra), LCL (logic calculi), NLP (natural language processing), SET (set theory), SWV (software verification), SYN (syntactic).
3. MML – *multi-modal logic*.
   20 problems in a multi-modal logic syntax from various textbooks and applications, e.g., security protocols and dialog systems [8, 21, 22].
4. NLP/SET – *classical logic*.
   80 problems from the NLP and SET domains of the TPTP library [23]; these allow comparisons with classical ATP systems.
5. SYM – *syntactic modal*.
   175 problems from various textbooks [9–12, 17, 20, 25] and 70 problems from the TANCS-2000 system competition for modal ATP systems [14].

3.2 Modal Problem Status and Difficulty Rating

As already done in the TPTP 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 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 other ATP system.

The *status* is either `Theorem`, `Non-Theorem` or `Unsolved`; see [18]. For problems with status `Theorem` (i.e. valid) or `Non-Theorem` (i.e. invalid) at least one of the considered ATP systems has found a proof or counter model, respectively. Problems with

[1] The domains MML, NLP, and SET were added in v1.1 of the QMLTP library.

Table 2. QMLTP library v1.1: Status and rating summary for all uni-modal problems

Logic	Domain Condition	Modal Status: Theorem	Modal Status: Non-Theorem	Modal Status: Unsolved	Modal Rating: 0.00	Modal Rating: 0.01–0.49	Modal Rating: 0.50–0.99	Modal Rating: 1.00	∑
K	varying	105	134	341	72	0	167	341	580
	cumulative	136	204	240	54	42	244	240	580
	constant	161	159	260	54	50	216	260	580
D	varying	180	243	157	65	46	312	157	580
	cumulative	200	257	123	53	74	330	123	580
	constant	219	242	119	55	90	316	119	580
T	varying	228	173	179	102	49	250	179	580
	cumulative	255	158	167	73	94	246	167	580
	constant	274	144	162	74	110	234	162	580
S4	varying	278	146	156	125	53	246	156	580
	cumulative	342	120	118	91	106	265	118	580
	constant	358	106	116	88	121	255	116	580
S5	varying	362	115	103	156	62	259	103	580
	cumulative	441	64	75	109	179	217	75	580
	constant	443	64	73	110	179	218	73	580

`Unsolved` status have not been solved by any ATP system. No inconsistencies between the output of the ATP systems were found. The status is specified with respect to a particular modal logic and a particular *domain condition*.

When determining the modal status, the standard semantics of first-order modal logics are considered; see e.g. [10]. 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. For release v1.1 of the QMLTP library all rating and status information is with respect to the first-order modal logics K, D, T, S4, or S5 with constant, cumulative or varying domain condition [10, 25].

To determine the modal rating and status of the problems, all existing ATP systems for first-order modal logic were used. These are LEO-II 1.2.6-M1.0, Satallax 2.2-M1.0, MleanSeP 1.2, MleanTAP 1.3, f2p-MSPASS 3.0 and MleanCoP 1.2. Not all systems support all modal logics or domain conditions. LEO-II [3] and Satallax [6] are ATP systems for typed higher-order logic. To deal with modal logic, both ATP systems use an embedding of quantified modal logic into simple type theory [2].[2] LEO-II uses an extensional higher-order resolution calculus. Satallax employs a complete ground tableau calculus for higher-order logic. MleanSeP and MleanTAP are compact ATP systems for several first-order modal logics.[3] MleanSeP is a compact implementation of the standard modal sequent calculi and performs an analytic proof search. MleanTAP implements an analytic free-variable prefixed tableau calculus and uses an additional prefix unification. f2p-MSPASS uses a non-clausal instance-based method

[2] LEO-II and Satallax were used as these are the two higher-order ATP system that solved the highest number of problems at CASC-23 and CASC-J5 [24].

[3] Available at `http://www.leancop.de/mleansep/` and `http://www.leancop.de/mleantap/`

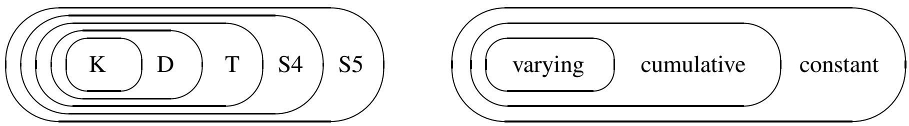

Fig. 1. The theoremhood relationship between different modal logics and domain conditions

and the MSPASS 3.0 system for propositional modal logic. MleanCoP implements a connection calculus for first-order modal logic extended by prefixes and a prefix unification algorithm [15]. Table 2 shows statistics about the status and rating of the problems in the QMLTP library for all uni-modal logics under consideration. The 20 problems for multi-modal logic contain 14 theorems and 6 unsolved problems. Figure 1 shows the theoremhood relationship, which is also reflected in Table 2.

3.3 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 [23] is extended by the modal operators. The two Prolog atoms "`#box`" and "`#dia`" are used for representing $\Box$ and $\Diamond$, respectively. The formulas $\Box F$ and $\Diamond F$ are then represented by "`#box:F`" and "`#dia:F`", respectively (see Figure 2). For multi-modal logic the modal operators $\Box_i$ and $\Diamond_i$ are represented by "`#box(i)`" and "`#dia(i)`", respectively, in which the index `i` is a Prolog atom. Furthermore, for multi-modal logic the `set_logic` command of the new TPTP *process instruction language* is used to specify the semantics of the used modal operators. For example,

```
tpi(1,set_logic,modal([(a,[s4,constant]),(b,[d,constant])])).
```

determines the specific semantics of the multi-modal operators $\Box_a$, $\Diamond_a$, $\Box_b$, and $\Diamond_b$.

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 first-order modal problem is shown in Figure 2.

3.4 Tools and Prover Database

The TPTP library provides the tptp2X tool for transforming and converting the syntax of TPTP problem files. This tool can be used for the QMLTP library as well. *Format files* for all existing modal ATP systems are included in the library. They are used together with the tptp2X tool to convert the problems in the QMLTP library into the input syntax of existing modal ATP systems. The prover database of the library provides information about published modal ATP systems. For each system some basic information is

```
%------------------------------------------------------------------------------
% File     : SYM001+1 : QMLTP v1.1
% 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   :      varying      cumulative   constant
%             K   Non-Theorem  Non-Theorem  Theorem       v1.1
%             D   Non-Theorem  Non-Theorem  Theorem       v1.1
%             T   Non-Theorem  Non-Theorem  Theorem       v1.1
%             S4  Non-Theorem  Non-Theorem  Theorem       v1.1
%             S5  Non-Theorem  Theorem      Theorem       v1.1
%
% Rating   :      varying      cumulative   constant
%             K   0.50         0.75         0.25          v1.1
%             D   0.75         0.83         0.17          v1.1
%             T   0.50         0.67         0.17          v1.1
%             S4  0.50         0.67         0.17          v1.1
%             S5  0.50         0.20         0.20          v1.1
%
%  term conditions for all terms:  designation: rigid,  extension: local
%
% Comments :
%------------------------------------------------------------------------------
qmf(con,conjecture,
(( ! [X] : (#box : ( f(X) ) ) ) => (#box : ( ! [X] : ( f(X) ) )))).
%------------------------------------------------------------------------------
```

Fig. 2. Problem file SYM001+1

provided, e.g., author, web page, short description, references, and test runs on two example problems. A summary and a detailed list of the performance results of the modal ATP system on the problems in the QMLTP library are included as well.

4 Conclusion

Extensive testing is an integral part of any software development. Logical problem libraries provide a platform for testing ATP systems. Hence, they are crucial for the development of correct and efficient ATP systems. Despite the fact that modal logics are considered as one of the most important non-classical logics, the implementation of ATP systems for (first-order) modal logic is still in its infancy.

The QMLTP library provides a comprehensive set of problems for testing ATP systems for first-order modal logic. Version 1.1 includes 600 problems with almost 9.000 status and rating information. 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. The availability of modal ATP systems that are sufficiently efficient will promote their employment within real applications. This will generate more modal problems from actual applications, which in turn will be included in the QMLTP library (provided that they will be submitted). The few multi-modal problems in the current release have already stimulated the implementation of the first ATP systems for first-order multi-modal logic. Future versions of the library will include more problems for multi-modal logic.

Like other problem libraries the QMLTP library is an ongoing project. All interested users are invited to submit new (first-order) modal problems and new modal ATP systems to the QMLTP library.

Acknowledgements. The authors would like to thank all people who have contributed to the QMLTP library so far. In particular, Geoff Sutcliffe and Christoph Benzmüller for many helpful discussions and suggestions.

References

1. Balsiger, P., Heuerding, A., Schwendimann, S.: A Benchmark Method for the Propositional Modal Logics K, KT, S4. Journal of Automated Reasoning 24, 297–317 (2000)
2. Benzmüller, C.E., Paulson, L.C: Quantified Multimodal Logics in Simple Type Theory. Seki Report SR-2009-02, Saarland University (2009) ISSN 1437-4447
3. Benzmüller, C.E., Paulson, L.C., Theiss, F., Fietzke, A.: LEO-II - A Cooperative Automatic Theorem Prover for Classical Higher-Order Logic (System Description). In: Armando, A., Baumgartner, P., Dowek, G. (eds.) IJCAR 2008. LNCS (LNAI), vol. 5195, pp. 162–170. Springer, Heidelberg (2008)
4. Blackburn, P., van Bentham, J., Wolter, F.: Handbook of Modal Logic. Elsevier, Amsterdam (2006)
5. Boeva, V., Ekenberg, L.: A Transition Logic for Schemata Conflicts. Data & Knowledge Engineering 51(3), 277–294 (2004)
6. Brown, C.E.: Reducing Higher-Order Theorem Proving to a Sequence of SAT Problems. In: Bjørner, N., Sofronie-Stokkermans, V. (eds.) CADE 2011. LNCS, vol. 6803, pp. 147–161. Springer, Heidelberg (2011)
7. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: EQL-Lite: Effective First-Order Query Processing in Description Logics. In: Veloso, M.M. (ed.) IJCAI 2007, pp. 274–279 (2007)
8. Fariñas del Cerro, L., Herzig, A., Longin, D., Rifi, O.: Belief Reconstruction in Cooperative Dialogues. In: Giunchiglia, F. (ed.) AIMSA 1998. LNCS (LNAI), vol. 1480, pp. 254–266. Springer, Heidelberg (1998)
9. Fitting, M.: Types, Tableaus, and Goedel's God. Kluwer, Amsterdam (2002)
10. Fitting, M., Mendelsohn, R.L.: First-Order Modal Logic. Kluwer, Amsterdam (1998)
11. Forbes, G.: Modern Logic. A Text in Elementary Symbolic Logic. OUP, Oxford (1994)
12. Girle, R.: Modal Logics and Philosophy. Acumen Publ. (2000)
13. Gödel, K.: An Interpretation of the Intuitionistic Sentential Logic. In: Hintikka, J. (ed.) The Philosophy of Mathematics, pp. 128–129. Oxford University Press, Oxford (1969)
14. Massacci, F., Donini, F.M.: Design and Results of TANCS-2000 Non-classical (Modal) Systems Comparison. In: Dyckhoff, R. (ed.) TABLEAUX 2000. LNCS (LNAI), vol. 1847, pp. 50–56. Springer, Heidelberg (2000)
15. Otten, J.: Implementing Connection Calculi for First-order Modal Logics. In: 9th International Workshop on the Implementation of Logics (IWIL 2012), Merida, Venezuela (2012)

16. Patel-Schneider, P.F., Sebastiani, R.: A New General Method to Generate Random Modal Formulae for Testing Decision Procedures. Journal of Articial Intelligence Research 18, 351–389 (2003)
17. Popcorn, S.: First Steps in Modal Logic. Cambridge University Press, Cambridge (1994)
18. Raths, T., Otten, J., Kreitz, C.: The ILTP Problem Library for Intuitionistic Logic. Journal of Automated Reasoning 38(1-3), 261–271 (2007)
19. Reiter, R.: What Should a Database Know? Journal of Logic Programming 14(1-2), 127–153 (1992)
20. Sider, T.: Logic for Philosophy. Oxford University Press, Oxford (2009)
21. Stone, M.: Abductive Planning With Sensing. In: AAAI 1998, Menlo Park, CA, pp. 631–636 (1998)
22. Stone, M.: Towards a Computational Account of Knowledge, Action and Inference in Instructions. Journal of Language and Computation 1, 231–246 (2000)
23. Sutcliffe, G.: The TPTP Problem Library and Associated Infrastructure: The FOF and CNF Parts, v3.5.0. Journal of Automated Reasoning 43(4), 337–362 (2009)
24. Sutcliffe, G.: The 5th IJCAR automated theorem proving system competition – CASC-J5. AI Communications 24(1), 75–89 (2011)
25. Thion, V., Cerrito, S., Cialdea Mayer, M.: A General Theorem Prover for Quantified Modal Logics. In: Egly, U., Fermüller, C. (eds.) TABLEAUX 2002. LNCS (LNAI), vol. 2381, pp. 266–280. Springer, Heidelberg (2002)