

The MMT API: A Generic MKM System

Florian Rabe

Computer Science, Jacobs University Bremen, Germany
<http://trac.kwarc.info/MMT>

Abstract. The MMT language has been developed as a scalable representation and interchange language for formal mathematical knowledge. It permits natural representations of the syntax and semantics of virtually all declarative languages while making MMT-based MKM services easy to implement. It is foundationally unconstrained and can be instantiated with specific formal languages.

The MMT API implements the MMT language along with multiple backends for persistent storage and frontends for machine and user access. Moreover, it implements a wide variety of MMT-based knowledge management services. The API and all services are generic and can be applied to any language represented in MMT. A plugin interface permits injecting syntactic and semantic idiosyncrasies of individual formal languages.

The MMT Language Content-oriented representation languages for mathematical knowledge are usually designed to focus on either of two goals: (i) the automation potential offered by mechanically verifiable representations, as pursued in semi-automated proof assistants like Isabelle and (ii) the universal applicability offered by a generic meta-language, as pursued in XML-based content markup languages like OMDOC. The MMT language [11] (Module system for Mathematical Theories) was designed to realize both goals in one coherent system. It uses a minimal number of primitives with a precise semantics chosen to permit natural and adequate representations of many individual languages.

A key feature is *foundation-independence*: MMT systematically avoids a commitment to a particular type theory or logic. Instead, it represents every formal system as an MMT theory: domain theories (like the theory **Group**), logics (like first-order logic FOL), and logical frameworks (like LF [4]) are represented uniformly as MMT *theories*. These theories are related by the *meta-theory* relation, e.g., LF is the meta-theory of FOL, which in turn is the meta-theory of **Group**. MMT uses this relation to obtain the semantics of a theory from that of its meta-theory; thus, an external semantics (called the *foundation*), e.g., a research article or an implementation, only has to be supplied for the topmost meta-theories. For example, a foundation for LF can be given in the form of a type system.

Theories contain typed *symbol declarations*, which permit the uniform representation of constants, functions, and predicates as well as – via the Curry-Howard correspondence – judgments, inference rules, axioms, and theorems. Theories are related via *theory morphisms*, which subsume translations, functors, and models. Finally, MMT provides a module system for building large theories and morphisms via reuse and inheritance.

Mathematical objects such as terms, types, formulas, and proofs are represented uniformly as OPENMATH objects [1], which are formed from the symbols available to the theory under consideration: For example, LF declares the symbols `type` and λ ; FOL declares \forall and \Rightarrow , and `Group` declares \circ and e . MMT is agnostic in the typing relation between these objects and instead delegates the resolution of typing judgments to the foundation. Then all MMT results are obtained for arbitrary foundations. For example, MMT guarantees that theory morphisms translate objects in a typing- and truth-preserving way, which is the crucial invariant permitting the reuse of results in large networks of theories.

The MMT API Exploiting the small number of primitives in MMT, the MMT API provides a comprehensive, scalable implementation of MMT itself and of MMT-based knowledge management (KM) services. The development is intentionally *application-independent*: It focuses on the data model of MMT and its KM services in a way that makes the integration into specific applications as easy as possible. But by itself, it provides only a basic user interface.

All algorithms are implemented generically and relegate all foundation-specific aspects to plugins. Concrete applications usually provide a few small plugins to customize the behavior to one specific foundation and a high-level component that connects the desired MMT services to a user interface.

The API is written in the functional and object-oriented language Scala [9], which is fully compatible with Java so that API plugins and applications can be written in either language. Excluding plugins and libraries, it comprises over 20,000 lines of Scala code compiling into about 3000 Java `class` files totaling about 5 MB of platform-independent bytecode. Sources, binaries, API documentation, and user manual are available at <http://trac.kwarc.info/MMT>.

Knowledge Management Services The MMT API provides a suite of coherently integrated KM services, which we only summarize here because they have been presented individually. A *notation language* based on [7] is used to serialize MMT in arbitrary output formats. Notations are grouped into *styles*, and a rendering engine presents any MMT concept according to the chosen style.

MMT content can be organized in *archives* [5], a light-weight project abstraction to integrate source files, content markup, narrative structure, notation indices, and RDF-style relational indices. Archives can be built, indexed, and browsed, and simplify distribution and reuse. A *query language* [10] integrates hierarchic, relational, and unification-based query paradigms. A *change management* infrastructure [6] permits detecting and propagating and changes at the level of individual OPENMATH objects.

User and System Interfaces If run as a standalone application, the API responds with a shell that interacts via standard input/output. The shell is scriptable, which permits users and application developers to initialize and configure it conveniently. For example, to check the theory *Group*, the initialization script would first register the MMT theory defining the syntax of LF and then a plugin

providing a foundation for LF, then register the theory FOL, and finally check the file containing the theory *Group*.

A second frontend is given by an HTTP server. For machine interaction, it exposes all API functionality and KM services via a RESTful interface, which permits developing MMT-based applications outside the Java/Scala world. For human interaction, the HTTP server offers an interactive *web browser* based on HTML+presentation MathML. The latter is computed on demand according to the style interactively selected by the user. Based on the JOBAD JavaScript library [3], user interaction is handled via JavaScript and Ajax. In particular, MMT includes a JOBAD module that provides interactive functionality such as definition lookup and type inference.

To facilitate distributing MMT content, all MMT declarations are referenced by canonical *logical identifiers* (the MMT URIs), and their physical locations (their URLs) remain transparent. This is implemented as a catalog that translates MMT URIs into URLs according to the registered knowledge repositories. MMT declarations are retrieved and loaded into memory transparently when needed so that storage and memory management are hidden from high-level services, applications, and users. Supported knowledge repositories are file systems, SVN working copies and repositories, and TNTBASE databases [12]. The latter also supports MMT-specific indexing and querying functions [8] permitting, e.g., the efficient retrieval of the dependency closure of an MMT knowledge item.

A Specific Application for a Specific Foundation The LATIN project [2] built an atlas of logics and related formal systems. The atlas is realized as an MMT project, and MMT is used for building and interactively browsing the atlas.

All theories in the atlas use LF as their meta-theory, which defines the abstract syntax of LF and thus of the logics in the atlas.

For concrete syntax, the Twelf implementation of LF is used. To integrate Twelf with MMT, LATIN developed an MMT plugin that calls Twelf to read individual source files and convert them to OMDOC, which MMT reads natively.

Based on this import, MMT's foundation-independent algorithms can index and catalog the LATIN atlas and make it accessible to KM services and applications. Here the use of Twelf remains fully transparent: An application sends only an MMT URI (e.g., the one LATIN defines for the theory FOL) to MMT and receives the corresponding Scala object.

From the perspective of MMT, Twelf is an external tool for parsing and type reconstruction that is applicable only to theories whose meta-theory is LF. From the perspective of Twelf on the other hand, the MMT theory LF does not exist. Instead, the symbols *type*, λ , etc. are implemented directly in Twelf's underlying programming language.

This is a typical situation: Generally, MMT uses the meta-theory to determine which plugin is applicable, and these plugins hard-code the semantics of the respective meta-theory. Similar concrete syntax plugins can be written for most languages and exist for, e.g., the ATP interface language TPTP, the ontology language OWL, and the Mizar language for formalized mathematics.

```

document derived.omdoc
remote module FalsityExt
remote module NEGExt
theory IMPExt meta lf
  include IMP
  imp2I : ((ded A → ded B → ded C) → ded A imp (B imp C))
         = [f:ded A → ded B → ded C]impI ([p:ded A]impI ([q:ded B]f p q))
  imp2E : (ded A imp (B imp C) → ded A → ded B → ded C)
         = [p:ded A imp (B imp C)][q:ded A][r:ded B]impE (impE p q) r
remote module CONJExt
remote module DISJExt
remote module Equiv

```

type

ded A imp (B imp C)

- infer type
- reconstructed types
- implicit arguments
- implicit binders
- redundant brackets
- Efold

LATIN also customizes the MMT web server by providing a style that provides notations for objects from theories with meta-theory LF. The above screenshot shows the generic web server displaying a theory `IMPExt`: It imports the theory `IMP` of the implication connective `imp` and extends it with derived rules for the introduction and elimination of double implications, i.e., formulas of the form $A \text{ imp } (B \text{ imp } C)$. The symbol `impI` represents the derivation of the rule $\frac{A, B \vdash C}{\vdash A \text{ imp } (B \text{ imp } C)}$. Via the context menu, the user has called type inference on the selected subobject, which opened a dialog showing the *dynamically inferred type*.

The interactive type inference is implemented using the HTTP interface to the MMT API. First of all, the LATIN style is such that the rendered HTML includes parallel markup in the form of special attributes on the presentation MathML elements. JavaScript uses them to build an MMT query that is posted to the server as an Ajax request and whose response is shown in the dialog. This query bundles multiple MMT API calls into a single HTTP request-response cycle: First the parallel markup is used to retrieve the OPENMATH object corresponding to the selected expression (and its context), then type inference is called, and finally the rendering engine is called to render the type as presentation MathML.

For the type inference, LATIN provides one further plugin: a foundation plugin that supplies the typing relation for theories with meta-theory LF. MMT uses it to perform type inference directly in memory without having to call external tools like Twelf. Such foundation plugins are easy to write because they can focus on the logical core of the type system and, e.g., parsing and module system remain transparent to the plugin. For example, the plugin for LF comprises only 200 lines of code

Except for the concrete syntax plugin, the presentation style, and the foundation plugin, all steps of the above example are foundation-independent and are immediately available for MMT content written in any other meta-theory. Moreover, being a logical framework, these plugins LF are immediately inherited by all logics defined in LF: We obtain, e.g., type inference for FOL and *Group* (in fact: for all logics defined in LATIN) without writing additional plugins.

Furthermore, all implementations are application-independent and can be immediately integrated into any application, e.g., a Wiki containing LF objects. This customization of MMT to specific foundations and specific applications occurs at minimal cost, a principle we call *rapid prototyping for formal systems*.

Acknowledgements Over the last 6 years, contributions to the API or to individual plugins have been made by Maria Alecu, Alin Iacob, Catalin David, Stefania Dumbrava, Dimitar Misev, Fulya Horozal, Füsün Horozal, Mihnea Iancu, Felix Mance, and Vladimir Zamdzhev. Some of them were partially supported by DFG grant KO-2428/9-1.

References

1. S. Buswell, O. Caprotti, D. Carlisle, M. Dewar, M. Gaetano, and M. Kohlhase. The Open Math Standard, Version 2.0. Technical report, The Open Math Society, 2004. See <http://www.openmath.org/standard/om20>.
2. M. Codescu, F. Horozal, M. Kohlhase, T. Mossakowski, and F. Rabe. Project Abstract: Logic Atlas and Integrator (LATIN). In J. Davenport, W. Farmer, F. Rabe, and J. Urban, editors, *Intelligent Computer Mathematics*, pages 289–291. Springer, 2011.
3. J. Gičeva, C. Lange, and F. Rabe. Integrating Web Services into Active Mathematical Documents. In J. Carette, L. Dixon, C. Sacerdoti Coen, and S. Watt, editors, *Intelligent Computer Mathematics*, pages 279–293. Springer, 2009.
4. R. Harper, F. Honsell, and G. Plotkin. A framework for defining logics. *Journal of the Association for Computing Machinery*, 40(1):143–184, 1993.
5. F. Horozal, A. Iacob, C. Jucovski, M. Kohlhase, and F. Rabe. Combining Source, Content, Presentation, Narration, and Relational Representation. In J. Davenport, W. Farmer, F. Rabe, and J. Urban, editors, *Intelligent Computer Mathematics*, pages 212–227. Springer, 2011.
6. M. Iancu and F. Rabe. Management of Change in Declarative Languages. In J. Campbell, J. Carette, G. Dos Reis, J. Jeuring, P. Sojka, V. Sorge, and M. Wenzel, editors, *Intelligent Computer Mathematics*, pages 325–340. Springer, 2012.
7. M. Kohlhase, C. Müller, and F. Rabe. Notations for Living Mathematical Documents. In S. Autexier, J. Campbell, J. Rubio, V. Sorge, M. Suzuki, and F. Wiedijk, editors, *Mathematical Knowledge Management*, pages 504–519. Springer, 2008.
8. M. Kohlhase, F. Rabe, and V. Zholudev. Towards MKM in the Large: Modular Representation and Scalable Software Architecture. In S. Autexier, J. Calmet, D. Delahaye, P. Ion, L. Rideau, R. Rioboo, and A. Sexton, editors, *Intelligent Computer Mathematics*, pages 370–384. Springer, 2010.
9. M. Odersky, L. Spoon, and B. Venners. *Programming in Scala*. artima, 2007.
10. F. Rabe. A Query Language for Formal Mathematical Libraries. In J. Campbell, J. Carette, G. Dos Reis, J. Jeuring, P. Sojka, V. Sorge, and M. Wenzel, editors, *Intelligent Computer Mathematics*, pages 142–157. Springer, 2012.
11. F. Rabe and M. Kohlhase. A Scalable Module System. *Information and Computation*, 2013. conditionally accepted; see <http://arxiv.org/abs/1105.0548>.
12. V. Zholudev and M. Kohlhase. TNTBase: a Versioned Storage for XML. In *Proceedings of Balisage: The Markup Conference 2009*, volume 3 of *Balisage Series on Markup Technologies*. Mulberry Technologies, Inc., 2009.