Representing, Archiving, and Searching the Space of Mathematical Knowledge

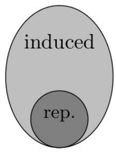
Mihnea Iancu, Michael Kohlhase, and Corneliu Prodescu

Jacobs University Bremen, Germany
http://kwarc.info

1 Introduction

There is an interesting duality between the forms and extents of mathematical knowledge that is verbally expressed (published in articles, scribbled on blackboards, or presented in talks/discussions) and the forms that are needed to successfully extend and apply mathematics. To "do mathematics", we need to judge the veracity, extract the relevant structures, and reconcile them with the context of our existing knowledge – recognizing parts as already known and identifying those that are new to us. In this process we may abstract from syntactic differences, and even employ interpretations via non-trivial mappings as long as they are meaning-preserving.

This mathematical practice of viewing an object of class A as one of class B – which we call **framing** – is an essential part of **mathematical literacy** – the skillset that identifies mathematical training. Indeed, framing is at the heart of understanding – seeing the network structure of math knowledge – and applying it. The essence of mathematical literacy is depicted in the figure on the right: trained mathematicians have access to a large, structured space of knowledge – we call it the **Mathematical Knowledge Space** (MKS) – that is induced via framing from a small core of represented knowledge. Unfortunately, mathematical software systems currently show only a very small degree of mathematical literacy. In this paper we present MMT theory graphs as a modular representation paradigm for mathematical knowledge, MathHub.info as an archive system that supports MMT-encoded knowledge, and $\flat$Search as an example of a math-literate search engine.



2 Representing the Math Knowledge Space in MMT

We will now present the OMDoc/MMT format [5] which focuses on the network structure of mathematical knowledge and makes framing a central representational concern: MMT groups symbols facts into **theories** and represents (potential) framings as *theory morphisms*, which interlink theories into a **theory graph**. Theory morphisms are mappings between theories which map axioms of the source theory to theorems of the target theory. This ensures that all theorems of the source theory induce theorems of the target theory.

H. Hong and C. Yap (Eds.): ICMS 2014, LNCS 8592, pp. 26–30, 2014.

To understand the setup, consider the theory graph in Fig. 1. The right side of the graph introduces the elementary algebraic hierarchy building up algebraic structures step by step up to rings; the left side contains a construction of the integers. In this graph, the nodes are *theories*, the solid edges are *imports* and the dashed edges are *views*.

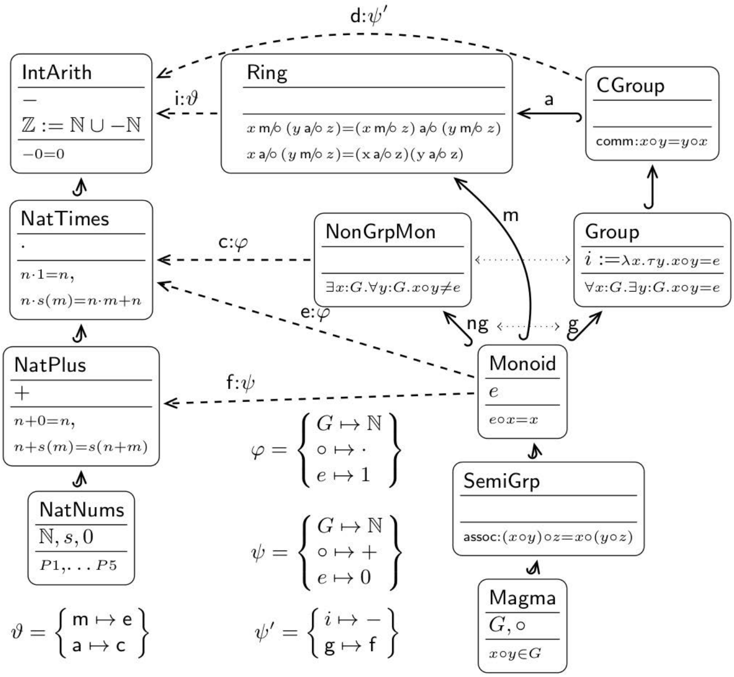


Fig. 1. A MMT Graph for Elementary Algebra

Importantly, every MMT symbol and statement is identified by a canonical, globally unique URI (called its **MMT URI**). Theories and views can be referenced relative to the URI of the containing theory graph, and symbol declarations by the URI of the containing theory, separated by ?. For instance, if U is the URI of the theory graph in Fig. 1 then the theory NatPlus and its symbol + have URIs U?NatPlus and, respectively, U?NatPlus?+.

Theory inheritance is realized by *structures*, which are named imports (and defined using theory morphisms). *Includes* are trivial structures which are unnamed and total. Symbol declarations induced by structures and views can be referenced relative to their name, separated by /. For instance, the addition operation from Ring can be referenced with U?Ring?a/$\circ$.

The definition of the theory Ring makes use of two MMT structures: m (for the multiplicative operations) and a (for the additive operations). To complete the ring we only need to add the two distributivity axioms in the inherited operators m/◦ and a/◦. Furthermore, a theory morphism, f, is used to represent that natural numbers with addition (NatPlus) form a monoid (Monoid).

It is a special feature of MMT that assignments can also map morphisms into the source theory to morphisms into the target theory. We use this to specify the morphism c modularly (in particular, this allows to re-use the proofs from e and c). Note that already in this small graph, there are a lot of induced statements. For instance, the associativity axiom is inherited seven times (via inclusions; twice into Ring) and induced four times (via views; twice each into NatArith and IntArith). All in all, we have more than an hundred induced statements from the axioms alone. If we assume just 5 theorems proven per theory (a rather conservative estimation), then we obtain a number of induced statements that is an order of magnitude higher.

3 Archiving the Math Knowledge Space in MathHub.info

The MathHub.info system [1] is a development environment for active mathematical documents and an archive for flexiformal mathematics.

The MathHub.info system has three main components (the detailed architecture is presented in Fig. 2):

– the GitLab repository manager as the versioned *data store* holding the source documents
– the MMT system [4] as the *semantic service provider* that imports the source documents and provides services for them
– and the Drupal CMS as the *frontend* that makes the sources and the semantic services available to users.

Currently, the MathHub.info data store contains the following libraries of various degrees of formality:

– the SMGloM termbase with ca. 1500 small sTeX files containing definitions of mathematical terminology and notation definitions.
– ca. 6500 files with sTeX-encoded teaching materials (slides, course notes, problems, and solutions) in Computer Science,
– the LATIN logic atlas with ca. 1000 meta-theories and logic morphisms,
– the Mizar Mathematical Library of ca. 1000 articles with ca. 50.000 theorems, definitions, and proofs, and
– a part of the HOL Light Library with 22 theories and over 2800 declarations.

We have MMT importers for all MathHub.info libraries and, therefore, MMT services become available for them. Current services including HTML presentation, querying, type checking and change management.

On the frontend side, Drupal natively supplies uniform theming, user management, discussion forums, etc. We extend it with dedicated modules to connect with the source documents in GitLab (for editing) as well as the imports in MMT (for MMT services, e.g. HTML presentation). Moreover, the JavaScript library JOBAD makes the documents active by interfacing with MMT services to enable complex in-browser interactions.

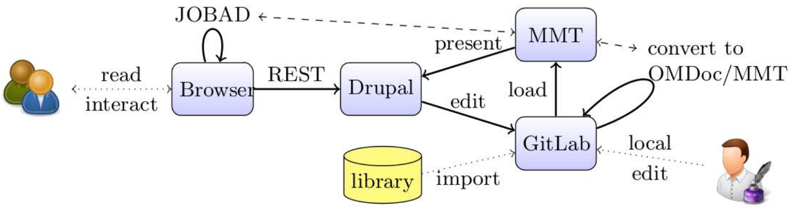


Fig. 2. The MathHub.info Architecture

4 Searching for Induced Statements

To search for induced statements, we use our MathWebSearch system [3], which indexes formula-URL pairs and provides a web interface querying the formula index via unification. This can be used for

Instance Search e.g. to find all instance of associativity we can issue the query $\forall x, y, z : \boxed{\mathrm{S}}.(x\boxed{\mathrm{op}}y)\boxed{\mathrm{op}}z = x\boxed{\mathrm{op}}(y\boxed{\mathrm{op}}z)$, where the $\boxed{\text{-}}$ are query variables that can be instantiated in the query. In the library from Fig. 1 we would find the commutativity axiom SemiGrp/assoc, its directly inherited versions in Monoid, Ring and in particular the version u?IntArith?c/g/assoc.

Applicable Theorem Search where universal variables in the index can be instantiated as well; this was introduced for a non-modular formal libraries in [2]. Here we could search for $3 + 4 = \boxed{\mathrm{R}}$ and find the induced statement u?IntArith?c/comm with the substitution $R \mapsto 4 + 3$, which allows the user to instantiate the query and obtain the equation $3 + 4 = 4 + 3$ together with the justification u?IntArith?c/comm that can directly be used in a proof.

Realizing ♭Search on top of MathWebSearch has two parts:

- The search engine proper is very simple: instead of harvesting formulae directly from a formal digital library, we flatten the library first, and then harvest formulae. Flattening is the process of explicating all induced statements in an OMDoc/MMT theory graph, a central service of the MMT system, and defining feature of the ♭Search system. Note that the MMT URIs of statements do not change during flattening, so they can directly be utilized as search hits in ♭Search.
- For the presentation of search hits, we cannot simply rely on the MMT system to dereference the MMT URIs (which would indeed compute the induced statements), but we have to use the structure of the OMDoc/MMT theory graph to explain the path between the search hits and the represented knowledge. Luckily the MMT URIs contain enough information to compute this. Fig. 3 shows a ♭Search result in action: ♭Search found the induced statement of associativity of $+$ on $\mathbb{Z}$ and uses the combinations of morphisms m and i from Fig. 1 to justify the hit.

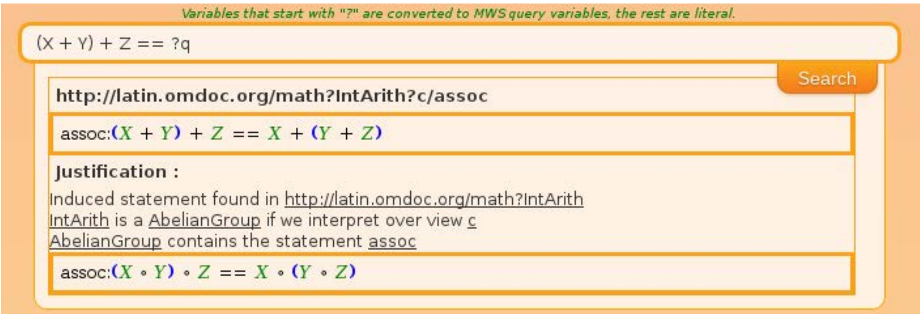


Fig. 3. The ♭Search Web Interface for MathHub.info

5 Conclusion and Future Work

We have presented a unified framework for representing the inherent network structure of mathematical knowledge – OMDoc/MMT –, for enabling mathematically literate services – MathHub.info – and substantiated this with a model service – ♭Search. The OMDoc/MMT language has been validated in large-scale representation and translation experiments, the systems are in a late prototype state; ♭Search is fully integrated into MathHub.info and can be used on the MathHub.info content directly (though results depend on the modular structure). We expect to open MathHub.info for general use in this year, when the system has stabilized.

References

1. Iancu, M., Jucovschi, C., Kohlhase, M., Wiesing, T.: System description: MathHub.info. In: Watt, S., Davenport, J., Sexton, A., Sojka, P., Urban, J. (eds.) CICM 2014. LNCS, vol. 8543, pp. 431–434. Springer, Heidelberg (2014)
2. Iancu, M., Kohlhase, M., Rabe, F., Urban, J.: The Mizar mathematical library in OMDoc: Translation and applications. Journal of Automated Reasoning 50(2), 191–202 (2013)
3. MathWebSearch searching mathematics on the web
4. Rabe, F.: MMT – a module system for mathematical theories
5. Rabe, F., Kohlhase, M.: Information & Computation (230), 1–54 (2013)