

Metadata for a wiki of formalized mathematics

Jesse Alama*

Center for Artificial Intelligence

New University of Lisbon

j.alama@fct.unl.pt

Abstract

In recent years wikis for formal mathematics have appeared. Formal mathematics presents a number of challenges for the wiki perspective. To enhance the quality of the data in these wikis from the perspective of information architecture, we propose some extensions of existing formal mathematics wikis to more properly handle *metadata*.

1 Introduction

Recent years have seen some attention paid to the problem of building wikis for formalized mathematical texts, and there are various proposals and even some live wiki or wiki-like systems for formal mathematics [9, 6]. Formal mathematics (or indeed any formally verified content) presents a number of challenges for the wiki perspective.

In this paper we propose some extensions of existing formal mathematics wikis to more properly handle *metadata*. The aim, ultimately, is to better apply the tools of *information architecture* [7] for a formal mathematical wiki. We sketch some salient kinds of metadata for formal mathematics and their implementation. Our work is but a modest step toward the grand aim of a semantic web in the domain of mathematics, with a bias toward formal verification.

2 Previous work

Our implementation concerns mizar, primarily, though our proposals naturally extend to other interactive theorem provers and indeed other formal wikis as well.

mizar is an interactive theorem prover built on classical first-order logic and set theory. Its associated library, the mizar Mathematical Library (MML), is a large corpus of formalized mathematical knowledge spanning many areas of mathematics.

One early resource is the mizar wiki, at <http://wiki.mizar.org>. This is a long-standing resource for those interested in the mizar language and its library. Part of that wiki documents metadata for the articles of the MML. Of interest to us are the tags assigned to articles from the American Mathematical Association's Mathematics Subject (MSC) Classification scheme.

However, the mizar wiki does not support editing of other metadata about articles of the library, and it serves its (HTML) resources without contentful metadata, neither in the HTTP headers nor in the representations themselves. The first wiki system for mizar is live at <http://mws.cs.ru.nl/mwiki>. This wiki does support editing mizar texts and ensures their validity. But this wiki also offers up only scant metadata. We build on earlier wiki efforts for formalized mathematics [9, 2], and on a dynamic website [1] for presenting fine-grained logical and mathematical dependencies for the mizar Mathematical Library. Our aim is to extend the infrastructure provided by these systems with rich metadata, thereby increasing the quality of mizar information available to humans and machines alike.

*The author was supported by the FCT project "Dialogical Foundations of Semantics" (DiFoS) in the ESF EuroCoRes programme LogICCC (FCT LogICCC/0001/2007).

The efforts of Kohlhase and his collaborators (e.g., [6]) toward semantic markup of mathematics, though not about formalized mathematics in the same sense as mizar texts, have helped to show ways of marking up mathematical texts.

3 Metadata categories

Some metadata about mizar texts are functions of the text—we can compute the metadata in question—and thus aren't suitable for user editing. The length of a mizar article in bytes, for example, or the underlying XML representation of an article, are properties of the text that cannot (so far as we can see) be meaningfully modified.

Other properties of texts, however, can be profitably exposed. The following properties of mizar texts are candidates for user-editable properties:

- Mathematics Subject Classification

Already there has been some informal efforts to annotate parts of the mizar Mathematical Library by assigning tags to articles from the American Mathematical Society's widely used Mathematical Subject Classification (MSC) scheme [4]. These data are used to seed our MSC assignments: we reuse the previous assignments of MSCs to mizar items (generally whole articles) for our initial assignment.

- Title and author information.
- Natural language abstract.
- Citations of relevant mathematical sources (e.g., books, papers, encyclopedia entries).
- Uncomputed or undetectable relationships between other items.

The mizar definition of cardinal numbers, for example, depends implicitly on the axiom of choice, but it is not clear how this can be detected from the dependency graph of the mizar Mathematical Library (if it is possible at all). It seems better to us to permit users of the mizar wiki to associate the mizar item that represents the definition of cardinal number with the mizar item that represents the axiom of choice. Such links could be represented through Atom link elements and served as See-Also HTTP headers.

- Alternates for theorems and definitions

Thus, a "named" theorem might have different variants or special cases. For example, the Jordan curve theorem can be understood in complete generality, or in the special case of polygons. A theorem might be stated for arbitrary dimensions, but be of especial interest for dimension 2 or 3 (e.g., Euler's polyhedron formula in full generality and the three-dimensional case). A theorem (or axiom, as the case may be) can even have different forms, such as Zermelo's well-ordering principle ("every set can be well-ordered") and a 'vanilla' rendering of the axiom of choice ("every non-empty set of non-empty sets has a choice function") are equivalent to one another (in a suitable background theory, of course).

Some of these metadata categories are already handled outside the context of our wiki. When authors submit their mizar texts to SUM (the mizar User Group, charged with maintaining mizar and its library), they supply the text as well as author and title information, an abstract (in English), and citations to relevant sources (e.g., the source of the "informal" proofs that are formalized in the article). This information is used to automatically generate a T_EX representation of an article, which is then published in the

journal *Formalized Mathematics*¹. The editors of *Formalized Mathematics* are charged with maintaining this information. The data is thus static, in the sense that one cannot edit the published title and abstract of a paper that has already been published in a journal.

A wiki that permits editing of metadata that already exists under the aegis of *Formalized Mathematics* is not intended to compete with that journal. The SUM and its efforts to professionalize the writing of formal mathematical texts (in mizar) complement the project proposed here. We use the mass of metadata that has been accumulated throughout the construction of the MML to seed our wiki with an initial pool of trustworthy data. Our aim is to make the sea of (mizar-)formalized mathematical knowledge discoverable and improvable. Where possible, we provide links to the authoritative, original sources for mizar formalizations.

Metadata support in the mizar language itself is (almost) entirely lacking. One cannot use the language itself to indicate, for example, the author of a mizar text. The language supports only one metadata-related construction, the `section` keyword, which takes no arguments and it intended to delimit different parts of an article.

In the context of a live wiki, however, curation-by-committee is not a viable option for eliciting and maintaining metadata. One must go beyond the mizar language and provide a means of attaching or editing metadata. Since one cannot directly add this information to mizar texts, in the context of a web interface, one must enter the information in a form.

The absence of metadata functionality from the mizar language (apart from the aforementioned `section` keyword) is, to some extent, an advantage, because it gives a clear separation of responsibility of the validity of the text from information about the text. Our wiki process for registering and maintaining metadata can complement the standard process for eliciting this data. With each new release of the mizar Mathematical Library, we receive another batch of “canonical” or official metadata for updated contents of the library.

4 Serving metadata

We serve metadata in two ways: in the HTTP headers, and in the body of our article representations. We adhere to RESTful principles [3], promoting the transparency of the HTTP protocol with the ultimate aim of making our (meta)data accessible to humans and machines (e.g., search engines) alike.

We use the Atom Publishing Protocol and its extensible notion of link [8] by serving metadata in the HTTP response headers. Our aim coheres with the goals of the Linked Open Data initiative². Given a request for, say, the 92nd theorem of the article POLYFORM, we can return the following headers:

```
# Request
GET /item/polyform/theorem/92 HTTP/1.1

# Response
HTTP/1.1 200 OK
Content-Type: text/html
Link: <http://wiki.example.org/item/polyform/theorem/92/text>;
    rel="alternate";
    type="text/plain"
Link: <http://wiki.example.org/authors/5>;
    rel="http://wiki.example.org/rels/author"

...
```

¹See <http://fm.mizar.org>.

²See <http://linkeddata.org/>

In this example, in addition to standard HTTP headers, we use the proposed Link HTTP header [5] to refer to the author of the article (if it is known). The representation in the . . . would likewise contain a link in its body to the author and the other metadata.

One challenge in the context of a wiki is to provide keep URIs “cool”, that is, essentially permanent objects. The slogan here is “Cool URIs don’t change” [3]. Metadata can help us to keep URIs cool while still admitting that some of the concepts, definitions, and theorems to which the URIs refer are evolving.

5 Outlook

Richer representations of formalized mathematical data and its metadata are available. Thus, one could serve RDFa or employ a suitable microdata framework. Richer representations of the mizar data itself as RDFa, or a documented, well-structured URI namespace for items, would be valuable. In the face of capable clients, one could even imagine performing non-trivial content negotiation; today, mizar text are served only as plain text or as (relatively thinly annotated) HTML.

Finally, it should be clear that although we focus on mizar, few parts of the project that directly depend on the mizar language, its tool set, or its library. The wiki principles developed here would apply to any other system as well. Indeed, a formal wiki for other systems, suitably equipped with the kind of metadata, would be a step toward setting up rich correspondences between various formal libraries and their web representations.

References

- [1] Jesse Alama. mizar-items: Computing fine-grained dependencies in the mizar mathematical library. To appear in *CICM 2011: Conference on Intelligent Computer Mathematics*.
- [2] Jesse Alama, Kasper Brink, Lionel Mamane, and Josef Urban. Large formal wikis: Issues and solutions. submitted, 2011.
- [3] Subbu Allamraju. *RESTful Web Services Cookbook*. O’Reilly, Sebastopol, California, 2010.
- [4] American Mathematical Society. *Mathematics Subject Classification*, 2010.
- [5] D. Connolly and I. Hickson. An entity header for linked resources. Technical report, Internet Engineering Task Force, April 1999. <http://www.w3.org/Protocols/9707-link-header.html>.
- [6] Christoph Lange and Michael Kohlhase. A semantic wiki for mathematical knowledge management. In Jörg Rech, Björn Decker, and Eric Ras, editors, *Emerging Technologies for Semantic Work Environments: Techniques, Methods, and Applications*, pages 47–68. IGI Global, April 2008.
- [7] Peter Morville and Louis Rosenfeld. *Information Architecture for the World Wide Web*. O’Reilly, Sebastopol, California, 3 edition, 2007.
- [8] The Atom publishing protocol. Available online at <http://tools.ietf.org/html/rfc5023>, October 2007.
- [9] Josef Urban, Jesse Alama, Piotr Rudnicki, and Herman Geuvers. A wiki for mizar: Motivation, considerations, and initial prototype. In Serge Autexier, Jacques Calmet, David Delahaye, Patrick D. F. Ion, Laurence Rideau, Renaud Rioboo, and Alan P. Sexton, editors, *Intelligent Computer Mathematics, 10th International Conference, AISC 2010, 17th Symposium, Calculemus 2010, and 9th International Conference, MKM 2010, Paris, France, July 5-10, 2010. Proceedings*, volume 6167 of *Lecture Notes in Computer Science*, pages 455–469. Springer, 2010.