

Metadata for a wiki of formalized mathematics

Jesse Alama

Center for Artificial Intelligence,
New University of Lisbon, Portugal

2011-08-27

Interest: a wiki made up of formal mathematical proofs

- My background: contribution to the MathWiki project at Radboud University Nijmegen.
- Initial system targeted at **mizar**.
- What is edited: **formal mathematical proofs**. (Not natural language.)
- Recently extended to the **CoRN** (Constructive **Coq** Repository at Nijmegen)
- Other libraries of formal mathematics are in our sights... (maybe you're next!)



Important missing piece: metadata

- Gap in current approach: **weak metadata support**.
 - Goal: richer representation-internal metadata.
 - Goal: augment the representation-internal metadata with representation-external metadata (HTTP headers).
 - Goal: support search for formalized mathematical content.



Current metadata: **mizar** wiki (<http://mws.cs.ru.nl/mwiki>)

HTTP header data for article FUNCT_1: 'Functions and their basic properties'

Date: Sat, 27 Aug 2011 04:57:43 GMT

Server: Apache/2.2.16 (Ubuntu)

Last-Modified: Sun, 03 Apr 2011 01:03:38 GMT

ETag: "3846b4-48a6d-49ff936912e80"

Accept-Ranges: bytes

Content-Length: 297581

Vary: Accept-Encoding

Content-Type: text/html



HTTP headers relevant for metadata

■ Standard headers

- **Pragma** (official HTTP header, but semantics not extensible and largely undefined)

■ Nonstandard headers

- **Link** (proposal since 1999; maybe some day it will be officially part of HTTP)
- **Resource-Description** (proposal coming from the RDF community)



Proposed use the (proposed) Link header

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"
```



Representation-internal metadata (**mizar**)

- title (contained in the XHTML `<title>` element, as well as in a semantically empty `` element)
- author (inside a ``)
- date received by the Association of **mizar** Users (inside a `` element)
- copyright information (inside a `` element)

(XHTML is badly invalid.)



Current metadata: CoRN wiki (<http://mws.cs.ru.nl/cwiki>)

Header information for <http://mws.cs.ru.nl/cwiki/metric2.Metric.html>

Date: Sat, 27 Aug 2011 05:06:16 GMT

Server: Apache/2.2.16 (Ubuntu)

Last-Modified: Fri, 13 May 2011 15:00:57 GMT

ETag: "620482-5880-4a32992b27c40"

Accept-Ranges: bytes

Content-Length: 22656

Vary: Accept-Encoding

Content-Type: text/html



Representation-internal metadata (Coq)

- No title, author, and copyright information (unlike the **mizar** example)
 - Copyright information **is** available, but it is lost by the **coqdoc** tool when transforming a proof script to an HTML representation.
- Comments from the Coq proof script are included (unlike the **mizar** example):

“In this version, a metric space over a setoid X is characterized by a ball relation B where $B\ e\ x\ y$ is intended to mean that the two points x and y are within e of each other ($d(x,y) \leq e$).”
- Divisions of the text wrapped in `<div>` elements, of the classes `code` and `text`

(XHTML is also badly invalid.)



Sources of relevant metadata

- Mathematics Subject Classification (MSC)
- Natural language abstract
- Citations of relevant mathematical sources (e.g., books or papers that the formalization follows or from which it diverges)
- Variants of a theorem or definition
 - Jordan curve theoremL complete generality vs. the case of polygons.
 - A theorem might be started for arbitrary dimensions, but be of especial interest for dimension 2 or 3.
 - Equivalent statements: Zermelo's well-ordering principle ('every set can be well-ordered'), 'every non-empty set of non-empty sets has a choice function')

Mathematics Subject Classification

- Product of the American Mathematical Society
- Large three-level hierarchy of mathematical topics (63 first-level categories, 400+ second-level categories, 5000+ leaf/third-level categories)
- Example:
 - First level: 20-XX Group theory and generalizations
 - Second level: 20Dxx Abstract finite groups
 - third level: 20D30 Series and lattices of subgroups

MSC for mizar

- Some assignment of MSC codes to **mizar** content **is already available**
- This information is not static; it is part of a wiki devoted to **mizar**.
- At present, the assignment of MSC codes (primary code, secondary code) is quite sparse: 273 out of 1132 articles (25%) have a primary MSC classification.

MSC for mizar (continued)

16D10 Modules, bimodules and ideals - General module theory

Primary classification

#	Article	
1	vectsp_2	Construction of Rings and Left-, Right-, and Bi-Modules over a Ring Michał Muzalewski
2	mod_1	Groups, Rings, Left- and Right-Modules Michał Muzalewski and Wojciech Skaba
3	lmod_5	Linear Independence in Left Module over Domain Michał Muzalewski and Wojciech Skaba
4	rmod_2	Submodules and Cosets of Submodules in Right Module over Associative Ring Michał Muzalewski and Wojciech Skaba
5	rmod_3	Operations on Submodules in Right Module over Associative Ring Michał Muzalewski and Wojciech Skaba
6	rmod_4	Linear Combinations in Right Module over Associative Ring Michał Muzalewski and Wojciech Skaba
7	rmod_5	Linear Independence in Right Module over Domain Michał Muzalewski and Wojciech Skaba
8	mod_2	Rings and Modules - Part II Michał Muzalewski
9	lmod_6	Submodules Michał Muzalewski
10	mod_4	Opposite Rings, Modules and their Morphisms Michał Muzalewski
11	lmod_7	Domains of Submodules, Join and Meet of Finite Sequences of Submodules and Quotient Modules Michał Muzalewski

MSC for **mizar**(continued)

- The **mizar** wiki at Nijmegen does not serve this information, nor is it editable (neither directly in **mizar** texts, nor as affiliated data).
- Presenting the information in a wiki form is a first step.
- The links to the articles are to **the MMLquery presentation of the article**, which is not editable.
- Links to the editable **mizar** wiki at Nijmegen would be preferable.

Future work

- Richer representations of formal mathematical texts is quite possible using. Lightweight (microformats) to heavyweight (RDFa) solutions are available.
- Associating metadata with formal mathematical texts can be done in various places. A natural candidate is to store the metadata in the content itself, using keywords that support such a practice.
 - At present, the format of **mizar** texts currently lacks such features.
 - Metadata is extracted solely from the disciplined placement of comments (thus for the **mizar** parser the data is not actually part of the text at all)
 - Other systems (e.g., Lange-Kohlhase infrastructure)

Literature

- Alama, J., Brink, K., Mamane, L., Urban, J.: 'Large formal wikis: Issues and solutions', CICM 2011.
- Allamraju, S.: **RESTful Web Services Cookbook**. O'Reilly, 2010.
- American Mathematical Society: **Mathematics Subject Classification (2010)**
- Lange, C., Kohlhase, M.: 'A semantic wiki for mathematical knowledge management', 2008.
- Morville, P., Rosenfeld, L.: **Information Architecture for the World Wide Web**. O'Reilly, 2007.
- **The Atom publishing protocol**
- **Reviving HTTP Header Linking: Some code and use cases**