

Mathematical Semantic Markup in a Wiki: the Roles of Symbols and Notations

3rd Semantic Wiki Workshop, ESWC 2008

Christoph Lange

Jacobs University, Bremen, Germany

KWARC – Knowledge Adaptation and Reasoning for Content

Digital Enterprise Research Institute, National University of Ireland, Galway

June 2, 2008

Overview

- Mathematical notation is surprisingly ambiguous and diverse. It's hard to make the semantics of a mathematical formula explicit.
- Semantic markup is a well-established solution – but not yet on the social semantic web!
- We've made OpenMath markup usable in a semantic wiki
- ... and gain as a benefit an improved editing of definitions of symbols and their notation.

Making Mathematical Wikis more Semantic

Why mathematics?

- foundational science for science, technology, and engineering

What does a wiki need to support mathematics semantically?

- \LaTeX formulæ often supported, e. g. in Wikipedia or PlanetMath
- What does $\mathcal{O}(n^2 + n)$ mean? – How many meanings can you identify?

Making Mathematical Wikis more Semantic

Why mathematics?

- foundational science for science, technology, and engineering

What does a wiki need to support mathematics semantically?

- \LaTeX formulæ often supported, e. g. in Wikipedia or PlanetMath
- What does $\mathcal{O}(n^2 + n)$ mean? – How many meanings can you identify?
 - 1 \mathcal{O} times $n^2 + n$
 - 2 function \mathcal{O} applied to $n^2 + n$
 - 3 Landau symbol: the set of all integer functions not growing faster than $n^2 + n$
- (3) is obvious?! – Let's *make* it obvious!

Content Dictionaries

How to make explicit that \mathcal{O} is a function set constructor?

- Define a vocabulary of mathematical symbols (cf. RDFS), a *Content Dictionary* (CD)
- E. g. a “function set” CD; agents can unambiguously refer to http://math-vocabulary.org/function-sets#Landau_0
- Formal CDs themselves contain mathematical formulæ, using previously defined symbols
- They are mathematical documents \Rightarrow make them collaboratively editable, too!

Semantic Markup with OpenMath

- OpenMath expresses the logical structure of a formula:
- Sublanguage for Content Dictionaries:
 - symbol name (\rightarrow URI)
 - informal description
 - additional properties (e. g. formal definition, axioms)
 - examples
 - other metadata (Dublin Core etc.)
 - human-readable notations (more about this later)
- Applications:
 - modular ontologies for data exchange between computer algebra systems or automated theorem provers (\rightarrow following talk)
 - Multilingual publishing
 - Adaptive learning applications
 - Web formula search engines
- OpenMath 3 to be aligned with MathML3 (W3C)

A Daunting Example

```

<CDDefinition>
  <Name>sin</Name>
  <Description>The sine function on real numbers</Description>
  <CMP>The sine function is defined in terms of the exponential function as
     $\sin x = \frac{1}{2}(e^{ix} - e^{-ix}).$ </CMP>
  <FMP>
    <OMOBJ xmlns="http://www.openmath.org/OpenMath"
      version="2.0" cdbase="http://www.openmath.org/cd">
      <OMA><OMS cd="relation1" name="eq"/>
        <OMA><OMS name="sin" cd="transc1"/>
          <OMV name="x"/></OMA>
        <OMA><OMS name="divide" cd="arith1"/>
          <OMA><OMS name="minus" cd="arith1"/>
            <OMA><OMS name="exp" cd="transc1"/>
              <OMA><OMS name="times" cd="arith1"/>
                <OMS name="i" cd="nums1"/>
                  <OMV name="x"/></OMA></OMA>
              <OMA><OMS name="exp" cd="transc1"/>
                <OMA><OMS name="times" cd="arith1"/>
                  <OMA><OMS name="unary_minus" cd="arith1"/>
                    <OMS name="i" cd="nums1"/></OMA>
                  <OMV name="x"/></OMA></OMA></OMA>
                <OMA><OMS name="times" cd="arith1"/>
                  <OMI>2</OMI>
                <OMS name="i" cd="nums1"/></OMA></OMA></OMA></OMOBJ></FMP>
    </CDDefinition>

```

Authoring, Presentation, and Navigation

- Semantic markup is hard to read and write \Rightarrow need authoring support, both for instances (formulæ) and vocabularies (CDs)
- Rendering semantic markup (e. g. to Presentation MathML or \LaTeX): need *notation definitions*
 - map symbols or patterns to fragments of presentation markup (MathML or \TeX)
 - recursively apply mappings for arguments of operators
- There is not the one single notation for a symbol: notations *change!*
 - time: $Rx \rightsquigarrow \sqrt{x}$ (16th century)
 - language: $\binom{n}{k} = C_n^k = C_k^n$
domain, style, level of sophistication, . . .
- Editor must reflect such changes!
- Editor should link rendered symbols to their definitions

The Semantic Wiki SWiM

- SWiM = IkeWiki [Schaffert] + MathML/**OpenMath**/OMDoc
- Knowledge representation “one page $\hat{=}$ one RDF resource” works well, but requires small pages (*Discuss!*)
- Page granularity found suitable for OpenMath: symbol definitions, formal/informal properties, examples, per-symbol and notation definitions
- OpenMath specification is different, considers a whole CD (mostly > 10 symbols) one “review unit”
- SWiM preserves compatibility by splitting documents into xi:included fragments on import
- Editor: semi-WYSIWYG, tabled XML structures, visual formula editor being integrated

The Semantic Wiki SWiM



User

- User Page (Administrator)
- Preferences
- Logout
- Theme: [tundra] [soria]

Navigation

- IkeWiki Help
- Recent Changes

Search

Edit

- Create Resource
- Create Class
- Create Property
- Create Multimedia
- Create Template
- Delete Resource
- Add Relation
- Remove Relation

System

Article [Discuss](#) [Metadata](#) [Context](#) [Edit](#) [Annotate](#) [History](#)

Languages: [\[en\]](#) [\[de\]](#) [\[pl\]](#) [\[bg\]](#) [\[it\]](#)

arith1

Identifier: cd:arith1

Types: [rdfs:Resource](#) - [omo:ContentDictionary](#) - [omo:OpenMathConcept](#) - [rdfs:Resource](#)

This CD defines symbols for common arithmetic functions. The symbol to represent the n-ary function to return the least common multiple of its arguments.

Symbol Definition (plus)

COMMENTED PROPERTY: for all a, b | $a + b = b + a$

FORMAL PROPERTY: $\forall a, b. (a + b) = (b + a)$

This symbol denotes unary minus, i.e. the additive inverse.

Symbol Definition (unary_minus)

COMMENTED PROPERTY: for all a | $a + (-a) = 0$

FORMAL PROPERTY: $\forall a. (a + -(a)) = 0$

The symbol representing a binary minus function. This is equivalent to adding the additive inverse.

Symbol Definition (minus)

COMMENTED PROPERTY: for all a, b | $a - b = a + (-b)$

FORMAL PROPERTY: $\forall a, b. (a - b) = (a + -(b))$

The symbol representing an n-ary multiplication function.

Symbol Definition (times)

EXAMPLE: $\left(\begin{pmatrix} 12 \\ 34 \end{pmatrix} \begin{pmatrix} 56 \\ 78 \end{pmatrix} \right) = \begin{pmatrix} 1922 \\ 4350 \end{pmatrix}$

References

outgoing

- [IkeWiki: hasAuthor](#)
- [omo: contains](#)
 - [omo: containsSymbolD](#)
 - [arith1+abs](#)
 - [arith1+divide](#)
 - [arith1+gcd](#)
 - [arith1+lcm](#)
 - [arith1+minus](#)
 - [arith1+plus](#)
 - [arith1+power](#)
 - [arith1+product](#)
 - [arith1+root](#)
 - [arith1+sum](#)
 - [arith1+times](#)
 - [arith1+unary minus](#)
 - [rdf: type](#)
 - [omo: ContentDictio](#)
 - [omo: OpenMathCon](#)
 - [rdfs: Resource](#)
- incoming
 - untyped

Socialise

- [Digg this](#)
- [Post to del.icio.us](#)
- [Post to Furl](#)
- [Post to Magnolia](#)
- [Post to Yahoo](#)

SITY

Making CDs Semantic with an Ontology

The OpenMath CD language has a well-defined semantics, but not in terms of the semantic web

- We developed an ontology for relevant structures and relationships in OpenMath CDs (e. g. “CD includes symbol definition”, “formula uses symbol”)
- On saving or importing, we extract RDF from OpenMath

```
<CD xml:id="sample">
  <CDName>sample</CDName>
  <CDDate>2008-03-05</CDDate>
  <CDStatus>private</CDStatus>
  <Description>A sample CD</Description>
  <xi:include href="url/of/CDDefinition"/>
</CD>
```

→

```
<#sample>
  rdf:type omo:ContentDictionary ;
  dc:identifier "sample" ;
  dc:date 2008-03-05 ;
  omo:status "private" ;
  dc:description "A sample CD" ;
  omo:containsSymbolDefinition <url/of/CDDefinition> .
```

Put this RDF into a Jena triple store; currently used for:

- incoming/outgoing links, grouped by property
- graphical RDF browser
- assistance in manual annotation
- inline SPARQL queries

Defining Notations for Symbols

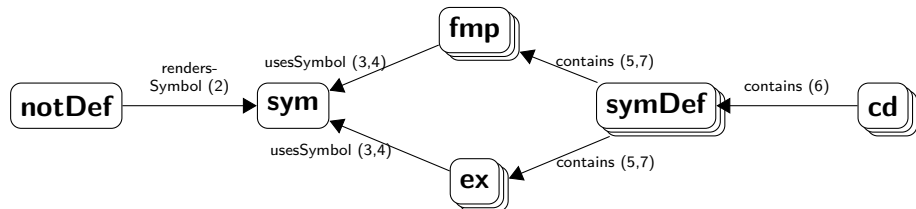
- Notation definitions map patterns of semantic markup to fragments of presentation markup, apply rules recursively
- e. g. $\textcircled{arith1\#root, \underline{arg}, \underline{n}} \vdash \sqrt[n]{\underline{arg}}$ (actually written in XML)
- link notation definition \rightarrow symbol definition in RDF:
`<#ntn-root> omo:rendersSymbol <url/of/arith1/root> .`
- What if an editor changes this?
 - 1 identify the change: XML diff
 - 2 identify documents affected by the change: query RDF graph for all resources using the symbol in a formula – and for all resources that xi:include the former

Determining the Effect of Changing a Notation Definition

```

1 SELECT DISTINCT ?page WHERE {
2   <changed-ntn-def> omo:rendersSymbol ?sym .
3   { ?page omo:usesSymbol ?sym } UNION
4   { ?exOrFMP omo:usesSymbol ?sym .
5     { ?page omo:contains ?exOrFMP } UNION
6     { ?page omo:contains ?symDef .
7       ?symDef omo:contains ?exOrFMP } } }

```



Related Work

- **Mathematical wikis:** not as semantic as SWiM, mostly using \LaTeX
- **Collaborative ontology editors:** OntoWiki (but more record-like structures)
- Other **editors for notation definitions:** $\text{PLAT}\Omega/\text{TEX}_{\text{MACS}}$, also reacts on changes, but not for web collaboration
- Rendering a formula is evaluating a (simple) **inline query** – a common feature in semantic wikis: “Find the notation definitions of the symbols used in the formula, present the results by rendering the formula.” (*Discuss!*)
- MONET **OpenMath ontology** for web service descriptions: one concept per standard symbol, e. g. `arith1#root`, but no CD structures

Conclusion and Outlook

Conclusion:

- Showed how semantic markup makes a semantic wiki aware of mathematics
- Integrated OpenMath into the semantic wiki SWiM:
 - 1 chose appropriate page granularity
 - 2 modeled ontology representing markup structures
 - 3 extracted relevant markup structures to RDF
- Used the semantics: Updates of notation definitions, navigation

Outlook (*Discuss!*):

- More complex reasoning tasks (transitive “CD uses CD” relationship): how to compute? Rules?
- Make domain knowledge available in semantic web compatible ontology (e. g. OWL-DL):

DifferentiableFunction \sqsubseteq ContinuousFunction