Bootstrapping a Semantic Wiki Application for Learning Mathematics

Claus Zinn*

1. Motivation

There is an enormous amount of math resources on the web: `http://www.mathcentre.ac.uk` (a British site), `http://www.mathe-online.at` (Austrian), `http://www.matheprisma.de` (German), and the virtual community site `http://planetmath.org`, to mention a few. Of course, our research group also created content, for instance, material for learning fractions, algebra, combinatorics, and statistics. Within the European project LeActiveMath, which aims at building an effective environment of technology-enhanced learning for mathematics (see `www.leactivemath.org`), we developed a significant amount of high quality teaching material in the math domain of differential calculus. The cost of authoring, however, has been high as our content providers were required to write content in a particular representation language, and to complement content with metadata along ontological and pedagogical dimensions. The extra effort, however, is not without reason as the resulting semantic-rich content is exploited by ActiveMath components for personalised course generation, semantic search, and problem solving support.

The rapid growth of math resources on the web, which is further pushed by wiki-based communities, is both a threat and an opportunity for intelligent math learning environments like ActiveMath. On the one hand, the continuing development of wiki content will soon dwarf ActiveMath's carefully authored semantic-rich content; and this will make its intelligent learning services that exploit such limited but high-quality content less enticing to use. On the other hand, math learning environments like ours could potentially profit from the availability of math resources on the web. If such content were to be enriched with machine-readable annotations that describe it, and if we could harness the collaborative authoring process and encourage and guide wiki authors to continually provide content and metadata, then intelligent services could unleash their true potential, with immediate return and added value for authors and learners.

In this paper, we describe se(ma)^{2}wi, a semantic math wiki application that we bootstrapped with high-quality content plus metadata from the LeActiveMath project. We anticipate that the availability of se(ma)^{2}wi, that is, the technical platform as well as the content resource, will help building and encouraging social communities to further make it grow, following, and potentially improving the blueprint of the original material.

*German Research Centre for Artificial Intelligence (DFKI), Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany, email: claus.zinn@dfki.de

2. The ActiveMath Learning Environment

The ActiveMath system is an environment for technology-enhanced learning [4]. It promotes the OMDoc representation language for writing mathematics.[1] This markup language extends OpenMath and MathML, and allows authors to represent the meaning of mathematical formulae instead of their presentation or appearance. It has also markup for the document and theory level of mathematical documents, providing authors with labels to categorise mathematical items in terms of “definition”, “assertion” (“theorem”, “lemma”), and “proof” as well as (pedagogically motivated) categories such as “note”, “exercise”, and “motivation”. To inform the course generator of ActiveMath there is also markup to label items in terms of “domain prerequisites”, “typical learning time”, “representation”, “field”, “learning context”, “competency level”, “abstractness”, and “trained competencies”.

ActiveMath authors use jEditOQMath, a purpose-built tool that supports the (template-based) creation, editing, validation and management of OMDoc documents for their use in the ActiveMath learning environment [2]. The resulting semantic-rich content enables semantic search, which also uses fuzzy matching [3], informs the personalised selection of mathematical content (course generation, given pedagogical strategies as well as individual user profiles and learner models) [5], provides interactive exercises with feedback [1], allows learners to access external tools (function plotter, concept map tool, computer algebra systems), and facilitates the multi-format presentation of math content using standard web browsers and document viewers.

To explore wiki technology and the potential to outsource content and metadata authoring to the wider community, we have built a semantic wiki with our semantic-rich content.

3. Wiki Technology

A website with wiki technology allows its visitors to view, add, remove, or otherwise edit and change (most of) its content with a standard web browser. The content is represented with a simple markup language; and it uses a subset of LaTeX for the authoring of math expressions. The high usability of the interface, the fact that changes take immediate effect, and the tools for discussing content and version control make a wiki an effective tool for collaborative authoring.

The Wikipedia website has a very large user and author base and enormous content that increases by the day, thus testifying the success of wiki technology. Wikipedia’s coverage on math topics is also considerable. There are, for instance, a number of articles with category “Differential_calculus”.[2] Yet, there are other wiki resources on differentiation. Wikibooks, the open-content textbooks collection, also features a book on calculus[3], whereas Wikiversity, a community for the creation and use of free learning materials and activities, has a School of Mathematics that offers, for instance, an introductory course to differentiation[4].

There are several interesting observations: Wikipedia exploits categories to automatically generate a page that lists all articles with a given category; Wikibooks adds a new granularity to wiki articles, namely books; and Wikiversity introduces the notion of mass customisation offering different courses for different target audiences (e.g., for pre-university and undergraduates).

[1] See `http://www.mathweb.org/omdoc`.
[2] See `http://en.wikipedia.org/wiki/Category:Differential_calculus`.
[3] See `http://en.wikibooks.org/wiki/Calculus`.
[4] See `http://en.wikiversity.org/wiki/Introduction_to_differentiation`.

The semantic annotation of this wiki math content is limited, however. Some articles have the category mark-up to link their page to an overall topic; but although pages are structured along explanations, definitions, proofs, and exercises, they are not labelled as such, and thus not open to machine interpretation. This is a pity since such annotations would facilitate the development and provision of intelligent services like semantic search (say, for all definitions within differential calculus), and personalised content selection and its presentation (say, the generation of additional examples or alternative proofs on request).

4. Porting Semantic Rich Content to a Semantic Wiki

We first installed MediaWiki[5], the software used for Wikipedia and related sites, and extended it with "semantic" functions (attributes and binary relations) from the ontoworld community.[6] This included the installation of texvc, a rendering engine for math formulae. To automate the import of articles, we also installed the python wikipedia bots package[7], and adapted some of its scripts.

```
<definition id="def_average_slope" for="diffquot_symbols/average_slope">
  <metadata>
    <Title xml:lang="en">Definition of the average slope between two points</Title>
    <extradata>
      <relation type="domain_prerequisite">
        <ref xref="lines_symbols/slope"/> <ref xref="lines_symbols/secant"/>
      </relation>
      <learningcontext value="secondary_education"/><learningcontext value="higher_education"/><learningcontext value="university_first_year"/>
      <field value="all"/>
      <typicallearningtime value="00:01:00"/>
      <representation value="verbal"/><representation value="visual"/><representation value="symbolic"/>
      <abstractness value="abstract"/>
    </extradata>
  </metadata>
  <CMP xml:lang="en">
    In order to determine the average slope of a curve between two points
    <OMOBJ> <OMA>
        <OMS cd="relation1" name="eq"/>
        <OMV name="P"/>
        <OMA> <OMS cd="basics_symbols" name="point_by_coordinates"/>
              <OMA> <OMS cd="basics_symbols" name="x_coord" />
                    <OMV name="P"/>
              </OMA>
              <OMA> <OMS cd="basics_symbols" name="y_coord"/>
                    <OMV name="P" />
              </OMA>
        </OMA>
      </OMA> </OMOBJ> and
    <OMOBJ>
      [...]
    </OMOBJ>,
    one replaces the curve by the <textref xref="lines_symbols/line">straight line</textref> through
    <OMOBJ>
```

⇓

```
'''average_slope'''
== Definition of the average slope between two points (def_average_slope) ==
[[is instance of::definition| ]]
In order to determine the average slope of a curve between two points
<math> {P}={\left({{x}_{{P}}},{{y}_{{P}}}\right)} </math>  and
<math> {Q}={\left({{x}_{{Q}}},{{y}_{{Q}}}\right)} </math> ,
one replaces the curve by the [[line|<nowiki>straight line</nowiki>]] through
<math> P </math>  and  <math> Q </math>
(i.e. by the [[secant|<nowiki>secant</nowiki>]] through  <math> P </math>  and  <math> Q </math> ).
[[Image:DefAverageSlopePQm.png]]

The average slope of the curve between  <math> P </math>  and  <math> Q </math>  is then defined to
be the [[slope|<nowiki>slope of this straight line</nowiki>]] <math> {P}{Q} </math> :
<br>
: <math> {{m}_{{P}{Q}}}={\frac{{{{y}_{{Q}}}-{{y}_{{P}}}}}{{{{x}_{{Q}}}-{{x}_{{P}}}}}} </math> .
<br>
'''Extradata'''
<br>
Needs as domain prerequisites [[slope]] [[secant]] <br>
[[learning context:=secondary_education| ]][[learning context:=higher_education| ]][[learning context:=university_first_year| ]]
[[field:=all| ]]
[[typicallearningtime:=00:01:00| ]]
[[representation:=verbal| ]][[representation:=visual| ]][[representation:=symbolic| ]]
[[abstractness:=abstract| ]]
```

Figure 1. XSLT conversion from OMDoc to wiki format

[5]See http://www.mediawiki.org.

[6]See http://wiki.ontoworld.org.

[7]See http://meta.wikimedia.org/wiki/Interwiki_bot/Getting_started.

The main task, however, has been the development and adaption of XSLT stylesheets that convert the XML-based OMDoc format into wiki-readable input. The first set of stylesheets were written to identify the input format in terms of the high-level categories such as “definition”, “theorem”, and “proof”. For each such segment, which resulted into a separate se(ma)^{2}wi article, we used a different set of stylesheets to translate its metadata into semantic relations and attributes, and to translate OpenMath representations of math expressions into LaTeX. Fig. 1 shows OMDoc input (a definition) and wiki output fragments of this translation. Note that the wiki article preserves most of the semantic information, except, for the time being, those of math formulae.[8] The example transformation yields

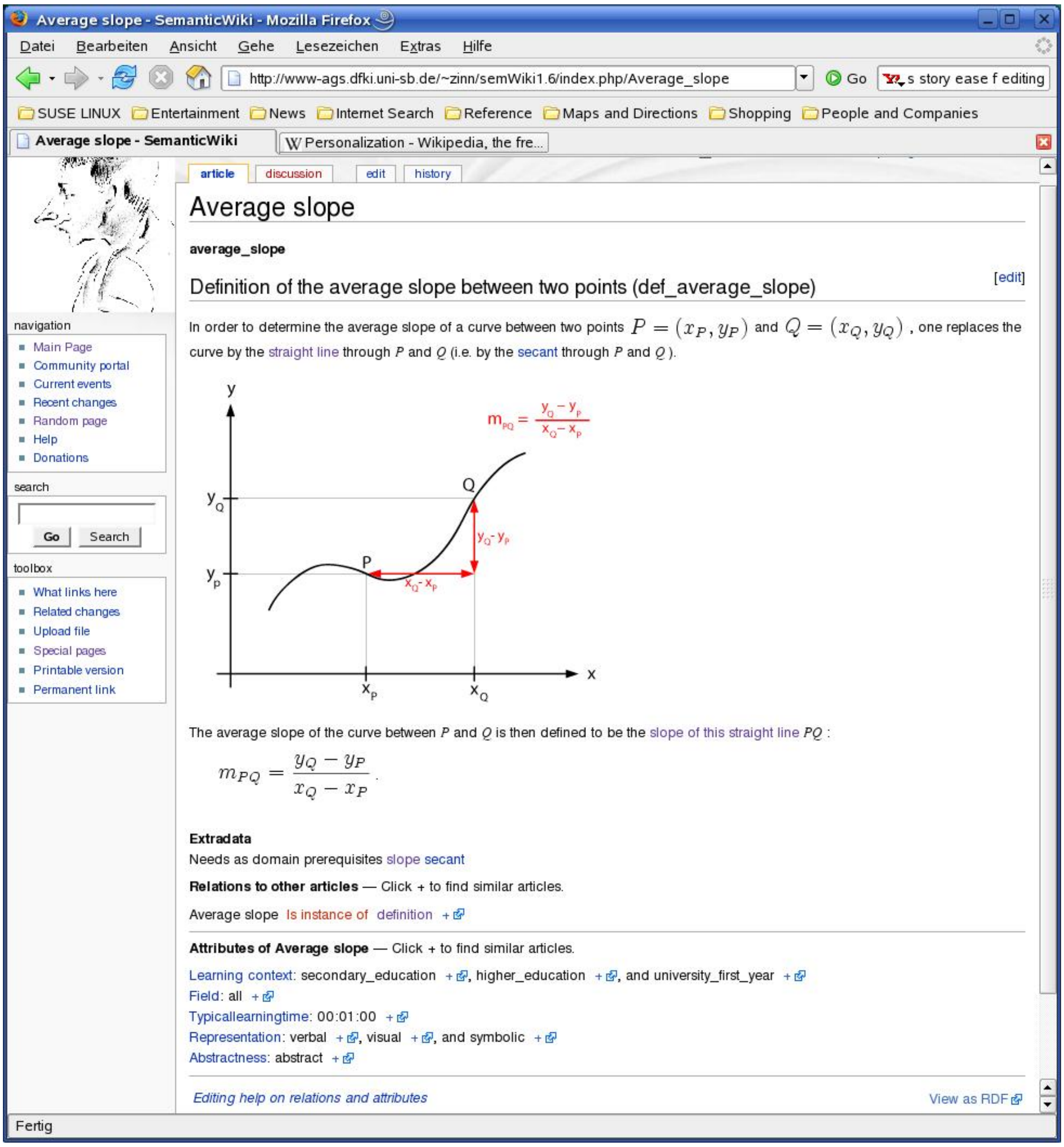

Figure 2. se(ma)^{2}wi **screenshot (page on average slope)**

a wiki article on the definition “average slope”, which MediaWiki renders with acceptable quality (see Fig.2). This page makes content references to the pages “straight line”, “secant” and “slope”. Below the content page, we see the metadata. The author’s annotations of “slope” and “secant” as domain prerequisite lead to the instantiation of the corresponding semantic attribute. Below, we see

[8] OMDoc segments that describe interactive elements were also not translated.

semantic information that specifies that this article is an instance of "definition", with several values for "learning context", "field", "typical learning time", "representation", and "abstractness". When a learner clicks on any of the attached links, the semantic wiki gives a list of all definitions, or, for instance, a list of all articles with attribute "learning time:=00:01:00". The available metadata thus offers effective navigational aids to browse through learning material.

In total, 1495 wiki articles have been imported from over 50 OMDoc source files. The resulting wiki articles include 121 definitions, 123 theorems, 35 lemmas, 208 proofs, 438 exercises (but without interactive elements), 74 notes, 47 introductions, and 405 examples.

5. Opportunities and Roadmap

We have imported all of the content that was authored within the LeActiveMath project into our semantic wiki. Before we start importing our other existing content, it is time for some reflection.

We were first surprised by the benefits of Wiki technology. One gets a simple, effective, and widely accepted content authoring tool, which also includes facilities for page preview, version control, and discussion. Content storage, indexing and retrieval is provided under the hood; and the presentation of content, especially the rendering of math expressions is satisfactory, at least. MediaWiki also delivers users and rights management; its semantic extension enables to search and organise wiki content via semantic relations. There is also editorial support via the wiki's special pages, where one can access, for instance, "long pages", "popular pages", and "dead end pages".

In ActiveMath, we had to implement those features ourselves, which, however, yielded more elaborated rendering options, a more sophisticated authoring environment, particularly for the provision of metadata, and more detailed semantic search facilities, to mention a few.

We are currently exploring the roadmap of future ActiveMath versions. Using our expertise, we could either adopt and improve semantic wiki technology as a pure authoring environment, or we could go the whole way: port our intelligent services, such as the course generation for personalised content selection, to an improved semantic wiki platform. In any case, the authoring community will need to agree on common standards for authoring mathematical content. Here, our se(ma)^{2}wi application, together with its semantic-rich content may serve as a technical vehicle and blueprint to promote ontological and pedagogical categories along the lines of the OMDoc standard.

Acknowledgements. se(ma)^{2}wi has been bootstrapped with content that has been developed by LeActiveMath project partners. The LeActiveMath project is funded under the 6th Framework Programme of the European Community (contract IST-2003-507826).

While ActiveMath has an open source license for academic use, the LeActiveMath content providers chose the adopt the Creative Commons License[9]: users are allowed to freely copy, distribute, display, and perform the content, given that they give credit to the original author; do not exploit the content for commercial purposes; and do not alter, transform, or build upon this content. We are currently investigating to change to a license that is more open to collaborative authoring, in particular, with respect to editing existing content.

[9] See `http://creativecommons.org/licenses/by-nc-nd/2.0/de/deed.en_GB`

References

[1] G. Goguadze, A. González Palomo, and E. Melis. Interactivity of exercises in ActiveMath. In C. K. Looi, D. Jonassen, and M. Ikeda, editors, *Towards Sustainable and Scalable Educational Innovations Informed by the Learning Sciences Sharing. Good Practices of Research Experimentation and Innovation*, volume 133 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, November 2005.

[2] P. Libbrecht and C. Gross. Experience report writing LeActiveMath calculus. In J. Borwein and W. Farmer, editors, *Proceedings of Mathematical Knowledge Management 2006*, number 4108 in LNAI. Springer Verlag, 2006.

[3] P. Libbrecht and E. Melis. Methods for access and retrieval of mathematical content in ActiveMath. In N. Takayama, A. Iglesias, and J. Gutierrez, editors, *Proceedings of ICMS-2006*, number 4151 in LNCS. Springer Verlag, 2006.

[4] E. Melis, G. Goguadze, M. Homik, P. Libbrecht, C. Ullrich, and S. Winterstein. Semantic-aware components and services of ActiveMath. *British Journal of Educational Technology*, 37(3), 2006.

[5] C. Ullrich. Tutorial planning: Adapting course generation to today's needs. In M. Grandbastien, editor, *Young Researcher Track Proceedings of 12th International Conference on Artificial Intelligence in Education*, pages 155–160, Amsterdam, The Netherlands, 2005.