How to make mathematical eContent travel well

Olga Caprotti, Matti Pauna, Mika Seppälä
University of Helsinki
{olga.caprotti, matti.pauna, mika.seppala}@helsinki.fi

Introduction

Electronic resources, that is eContent, produced for education ''travels well'' when it can be reused across national borders and in different curricular frameworks and, we may also add, across a variety of devices. Quality of the learning resource is naturally the prerequisite for it to be requested to travel: good eContent, if distributed on the web, will most likely be adopted outside of the course/school for which it was originally intended. This paper aims to discuss features of mathematical eContent that contribute to making it travel well. In particular, it focuses on various representation formats for mathematics, more specifically, on semantic markup such as Content MathML and OpenMath. Additionally, resources can be discovered on the web if they are correlated to metadata descriptions used by learning repositories.

This paper offers a brief survey into the technologies and activities of the JEM, Joining Educational Mathematics1, Thematic Network sponsored by the EU (Caprotti & Seppäla, 2008). Among the goals of JEM is the dissemination of best practice content for mathematics and the promotion of content enrichment activities of eContent in mathematics. Members of the JEM network include contributors to standard languages for the electronic communication of mathematics like OpenMath (Caprotti, Carlisle, & Cohen, 2000), MathML(Carlisle, Ion, & Miner, 2009), MathDox (Cohen, Cuypers, & Barreiro) and OMDOC(Kohlhase, 2006). Being a network of 20 nodes in 10 countries, it is natural that educational resources produced by the members have to travel well.

Learning Resources

Learning resources travel well, naturally, if they are high quality and in addition if they

- do not rely solely on written or spoken language
- avoid jargon used in country of origin
- are highly visual
- are modular
- clearly state the IPR.

Since mathematics is considered to be a universal language, one would expect any mathematics learning resource to travel well. However, not every format for writing

[1] http://www.jem-thematic.net

mathematics is created equal. The right choice of format for the representation of the mathematical objects can facilitate or impede adherence to some of these guidelines. Consider for instance the issue of notational conventions used in certain countries but not in others. If the mathematics is entered using a semantics representation, then notation can be generated depending on the preferences of the user of the document, either by location (intervals in France vs. USA) or by context (complex variables in engineering vs. physics). Moreover, the same mechanism extends also to the generation of formats that support alternative modalities of presentations such as aural or via assistive devices. Braille notation, which for mathematics is different in every country, can be more easily produced from a source in which the actual meaning of the mathematics is expressed. On the other hand, if the mathematics is represented only via images, there is very little one can do to change the notations or to run a screen reader or even to resize the formulae.

Below is a comparison chart to identify some of the characteristics of most relevant semantic markup formats now available to support mathematical content:

- Server: where and how it can be served
- Interactivity: if it can be used to represent interactive educational material like an exercise
- Notation support: if it allows users to set notation preferences (often this is also an aspect of the serving platform)
- Formulae: format in which the actual mathematics is encoded
- Authoring GUI: editing software
- Converters: alternative representations and formats which can be readily obtained from the sources

Format	Server	Interactivity	Notation support	Formulae	Authoring GUI	Representations
OMDoc	SWiM, pantha-rei, ActiveMath	Y	Y	OpenMath, C-MathML	STeX, OQMath, jEditOQMath	XHTML+MathM, LaTeX, PDF
Math-Dox	mathdox-player	Y	limited, via xslt	OpenMath, LaTeX, MathML	-	XHTML+MathML, PDF, LaTeX
CNXML	cnx.org	N	Y	C-MathML	LaTeX, Word	XHTML+MathML, PDF
OOXML	N	N	N	OMML	MS Office	PDF, PPTX, DOCX, DOC

The formats included in the chart are extensions of OpenMath or Content MathML such as OMDoc, MathDox, and CNXML(Henry, Baraniuk, & Kelty, 2003) and the Microsoft Office Open XML format called OOXML[2].

The choice of a specific format then depends on a number of factors, not least on the availability of conversion tools from legacy content and of convenient ways to edit material. It is clear that the richer the language, the more manipulation and processing one can expect to be able to do to adapt the material to arising requirements of presentation media, user preferences and learning resources. However, the authoring task today is still quite involved, especially when the document is able to support multiple views and functionality. Most of these formats still require expertise in XML and therefore their learning curve is steep. Here below we give a brief list of editors, mainly chosen to show the various possibilities for markup-based editing. This list is by no means exhaustive. The JEM portal devotes an entire section of the wiki to editing software[3], and also a few software description pages[4].

The **MathDox formula editor**, developed at TU/e, is a web-based editor for mathematical formulas that produces an OpenMath representation of the formula and can be easily integrated into interactive HTML pages, for instance to let users enter an expression.
Another OpenMath/MathML editor intended for web applications, and developed as a java applet, is the **WIRIS OpenMath tools**[5], available from Maths for More (Marques, Eixarch, Casanellas, & Martınez). For OpenMath advanced users, the applet can be configured to create new symbols, associate one or several graphical representations to it and, finally, use the new symbols through a toolbar in the editor.
The java editor **jEditOQMath** is supports the creation, editing and to some extent, the management of collections of OMDoc documents for use in the ActiveMath learning environment (Libbrecht & Gross). It allows XML-editing, validation, and template-supported creation of OQMath documents that can be automatically published on the ActiveMath server. The input of mathematical formulae is simplified by using a natural syntax such as $3a+5b$ or $Hn = set(x | x □ Rn □ π(n,x) > 0)$, then converted on-the-fly to OpenMath.

LaTeXML, by Bruce Miller, would be the choice for those authors already familiar with LaTeX, the widely used typesetting system for mathematics. Its stated goals are a faithful emulation of TeX's behavior, extensibility by packages, preservation of both semantic and presentation cues, abstract LaTeX-like, extensible, document type and possibility to support the semantics of mathematical content by good Presentation MathML, and eventually Content MathML and OpenMath. Similarly Jacobs University is developing **sTeX** (Kohlhase, 2006), Semantically Enhanced TeX, that provides specialized macro packages adding semantic information to a document without changing the visual appearance. This additional semantic information allows to convert to content-oriented, XML-based for-

---

2 http://www.ecma-international.org/publications/standards/Ecma-376.htm
3 http://www.jem-thematic.net/en/node/232
4 http://www.jem-thematic.net/view/software
5 http://www.wiris.com/demo/omeditor

mats, like CNX, Content MathML, OMDoc, Dublin-Core, and PhysML. Note that using this process on own legacy documents will involve to some extent the creation of new "LaTeXML bindings", namely specification of how to convert to the target representation in XML.

Learning repositories

Once produced, eContent is very valuable and can be shared at best by registration in a learning repository where it can be browsed or searched. The discovery of a learning resource, one that can be reused in the classroom, relies on the usage of appropriate keywords and of metadata standards, which describe its characteristics, including the IPR. The LOM, Learning Object Metadata, defines 70 fields in 9 categories aimed at specifying various aspects of a resource that might be useful when searching. A complete metadata record for a resource is entered usually by the author, e.g. using a web form provided by the repository, and filling in description of the resource type (exercise, simulation, lecture, graph, questionnaire, diagram, figure, index, slide, table, narrative text, exam, experiment, problem statement, self assessment), and interactivity type (active, expositive, mixed). However, because of the tediousness of the task, there are also efforts to produce and derive parts of the metadata automatically by online services that analyze the format of the resource and try to guess from it.

Within the JEM Thematic Network, we have developed a repository of learning resources that can be commented and reviewed by users and are meant to show the technologies developed by some of the partners of the network. The repository is conformant with the Open Archive Initiative and acts as a data provider that can be searched remotely by federated search engines. This ensures that resources registered at JEM have greater visibility and thus the potential to travel far.

Time travelling eContent

EContent does not only travel in space, across the web, but for any author, it travels in time. Courseware changes in time due to changes in the technology available to students and teachers alike, so, even if the actual content of the lecture is the same from one year to the next, the resources are delivered in a different way. We have gone from lecture notes to slides presentation, to podcast and now to mobile content. Re-incarnating resources in different formats is easier said than done. In our specific experience, the issues presented themselves with a collection of interactive exercises, developed over the years, by several authors and tailored to a variety of courses. The typical problems arising in electronic archiving and managing:

- no one knows anymore what material is there in its entirety
- multiple authors, multiple host locations, dependant on several versions
- non-uniform file naming scheme
- content tailored specifically for ad-hoc use cases.

The exercises were all developed as Maple TA exercises however the Maple TA administrative interface offers only limited support for the organization, search and browsing of existing collections. Classification and organization in question banks, each split in topics, is left to the single author.

Based on our experience, a reorganization that follows the guideline that the naming scheme for a question bank should encode a topic classification helps with later automatic processing and metadata generation of the exercises. For instance, exercises on fractions are collected in a question bank named *010202_Fractions* where 010202 is the classification ID in the Living Taxonomy for fractions. In the specific case of collections of exercises, LOM metadata can be generated semi-automatically since it is possible to default certain characteristics. More precisely the format is set to "application/mapleta", the interactivity type is "Active", the learning resource is "Exercise" and the intended end user role is "Learner". Further guidelines to make collections of exercises travel well in time are: topics should be kept small, one version only should be maintained and in order to locate exercises easily, they should be collected in a unique place.

To manage our collection, we have built a specific repository with enhanced browse and search facility and imported to it the entire reorganized collection of Maple TA question banks. This repository serves the need of authors, wanting to locate specific fragments of code, as well as lecturers, searching for exercises in a specific subject.

Conclusions

Even thought we often do not pay attention to the format we use for the electronic content we produce, our choice of a format over another will impact how well the material can travel to places, and through time. In general, the fact that the richer the format, the more involved the authoring process is, is especially true in the case of mathematics. The learning curve for producing semantic markup is steep however even today there are solutions that make the job possible. The advantage is the possibility to generate alterantive representations for the same content that adapt to the students' preferences in notation, to the device but also to the lecturers' needs.

Bibliography

Archambault, D., Fitzpatrick, D., Gupta, G., & Karshmer, A. (2004, Jan 1). Towards a universal maths conversion library. *Proc. ICCHP* .

Caprotti, O., & Seppäla, M. (2008, Sep 5). *Deliverable D1.3.2.* Retrieved from JEM Thematic Network: http://www.jem-thematic.net/en/D1.3.2

Caprotti, O., Carlisle, D., & Cohen, A. (2000, Jan 1). The OpenMath Standard. *The OpenMath Consortium* .

Carlisle, D., Ion, P., & Miner, R. (2009, Mar 6). Mathematical Markup Language (MathML) Version 3.0 . *W3C Working Draft* , 1-333.

Cohen, A., Cuypers, H., & Barreiro, E. (n.d.). MathDox: Mathematical Documents on the Web, Contribution to OMDoc book. *win.tue.nl* .

Henry, G., Baraniuk, R., & Kelty, C. (2003, Jan 1). The connexions project: Promoting open sharing of knowledge for education. *scholarship.rice.edu* .

Kohlhase, M. (2006, Jan 1). OMDoc--An Open Markup Format for Mathematical Documents [version 1.2]: Foreword by Alan Bundy (…. *portal.acm.org* .

Lange, C. (2008, Jan 1). Swim-a semantic wiki for mathematical knowledge management. *LECTURE NOTES IN COMPUTER SCIENCE* .

Libbrecht, P., & Gross, C. (n.d.). Authoring LeActiveMath Calculus Content. *Mathematical Knowledge Management* .

Marques, D., Eixarch, R., Casanellas, G., & Martınez, B. (n.d.). WIRIS OM Tools a Semantic Formula Editor. *icm2006.org* .