Early Examples of Software in Mathematical Knowledge Management

Patrick Ion[1,2]

[1] Mathematical Reviews [AMS], USA
ion@ams.org
[2] University of Michigan, USA
pion@umich.edu
http://www-personal.umich.edu/~pion/

Abstract. There are new roles for software in mathematical knowledge management (MKM). Three simple initial examples of MKM roles will be considered here. The first is software applied to the Mathematical Subject Classification (MSC). The second example is MathML (Mathematics Markup Language), a standard from the W3C, now in its third edition, and hoping to become an ISO standard. The third example of software in the service of mathematical knowledge is the use of programs to analyze the nature of our subject as represented by its literature seen as a network. How these tools have already been deployed makes clear that mathematical knowledge management, even in its primitive present form, can aid further development of mathematics. These examples above are just starting points.

Keywords: mathematical knowledge management, mathematical subject classification, MSC, mathematical markup language, MathML, SKOS, network analysis.

1 Introduction

Modern computers are often employed to do the calculations needed for mathematics, whether numerical or symbolic. It can be claimed that's why they were invented. There are also roles for software in mathematical knowledge management (MKM). An obvious one developing new tools to help access the mathematical literature, which is the main way mathematical knowledge has been expressed and archived until recently. Three simple initial examples of MKM roles will be considered here. The first is software applied to the Mathematical Subject Classification (MSC) which is a traditional way of organizing literature holdings.

2 Mathematics Subject Classification

A simple hierarchy of areas and sub-areas of mathematics is the basic structure of most traditional classifications of mathematics (or any other field) and, in particular of the Mathematics Subject Classification (MSC) developed in the 1960s

H. Hong and C. Yap (Eds.): ICMS 2014, LNCS 8592, pp. 31–35, 2014.

and now jointly maintained by Mathematical Reviews (MR) and Zentralblatt für Mathematik (ZB). Both these secondary knowledge services are probably now better known for their online databases MathSciNet [MathSciNet:website] and zbMATH [zbMATH:website].

In its present form the MSC skeleton is a rather flat three-level tree: 1 root, 63 top-level areas, 528 secondary areas, and 5606 leaf nodes. This gives an idea of its size. Over the years the MSC has been revised several times since the fields of mathematics vary in their importance and new views of mathematics and new concepts arise. Also mathematics can be said to have lost areas once regarded as within its purview, such as many of its classical applications, and especially statistics (applied statistics), and most recently computer science or informatics.

The most recent revision of the MSC resulted in the current MSC2010 which has its own web site, [MSC2010:website], used in preparing the revision during 2006–2009. Until next revision, MSC2020 — which I tend think of as the "Full Hindsight Revision", the web site will be the public archive of information about the MSC and deliberations about its development, and a proposed location for some services based on the MSC.

There are a several aspects of the details of the MSC that make for complications in realizing the MSC in software; for instance:

- It was developed from input by the highly varied mathematical community interpreted by, on the order of, a hundred mathematical editors from MR and ZB over decades.
- There are numerous additional relationships going beyond inclusion between areas identified.
- There is tension between the simple form of the MSC tree and faceting expressed in what has been developed by a heterogeneous collection of contributing authors.
- There is reuse of terminology in the node descriptions: linguistic overloading with mathematics.
- There are multilingual problems that arise from the international desire for translations of the MSC (e.g. Chinese, Russian, and Italian so far).
- The master MSC versions had been encoded since 1984, for the dominant purposes of printing them, using TEX typesetting system (which is admittedly a full macro computing language, in principle).
- There is a desire to record the evolution of the MSC over the years.
- It is not trivial to maintain an evolving labeled tree.
- There are mathematical formulas present in the descriptions.

The revision to MSC2010 was taken as a good time to change the authoritative source from the MSC to a form more promising for the Semantic Web. But it was also a time in which suggestions from the mathematical community could be collected through a web site, stored in a MySQL database so that all would be dealt with, and the changes being adopted could all be exposed with MediaWiki to public view.

MSC2010 information is now held in a master file encoded using SKOS (Simple Knowledge Organization System) [SKOS:website] which is a World Wide Web

Consortium standard [W3C:website]. Conversion to SKOS was done using Perl and Python scripts. But this eventually involved some small customizations of the vocabulary, which was envisaged by SKOS. Our needs went beyond the paradigms seen from the use by the US Library of Congress in converting their LC Subject Headings to SKOS (they have over 250,000 of those).

Of course, it is probably MR and ZB, and the traditional publishing world who make most use of the MSC in the course of their daily workflows. However, it does play a role in traditional searches for mathematical material, and can be used to make phrase-based searching more nuanced. To encourage the creation of more tools using the MSC the authoritative information is offered publicly in many forms: [MSC2010:SKOS], RDF/XML, Turtle, N-Triples, TriX, and JSON, as well as on-screen display versions in English, Chinese, Russian, Italian and the [MSC2010:MediaWiki] and the [MSC2010:TiddlyWiki]. There the beginnings of SPARQL access with prototype examples. In addition there are some classic text, TEX and PDF forms as well as a KWIC index on the site.

3 Mathematics Markup Language — MathML

One special aspect of the MSC was the inclusion of some mathematical formulas. Mathematical expressions are nowadays properly encoded for the web in MathML, our second example. MathML [MathML3:spec] is also a standard from the W3C, now in its third edition, and hoping to become an ISO standard soon.

MathML is a markup language for mathematical expressions. It was originally developed, starting in 1998, as an XML [XML:spec] vocabulary to support mathematical publication in the modern information world and was apparently oriented toward XHTML for the rest of the documents where formulas were to be found. As such MathML specifies a class of labeled rooted planar trees, but the details are significant. The changing web standards landscape and the rise of HTML5, an extensive rework of HTML, have shown that MathML can work in the new context with surprisingly little adjustment. The purpose of MathML is to capture both presentational aspects and some of the semantics of mathematics, so MathML is in the tradition of the efforts at pasigraphy reported at the first ICM in 1897 [Schröder:1897][Peano:1894], and also harks back to Leibniz's *calculus ratiocinator*. In its newest version 3.0 (Second edition) of 10 April 2014 MathML plays very well with HTML5. In turn HTML5 recognizes the `math` element from the MathML namespace and specifies that the semantics of markup within that element shall be defined by the MathML specification (and other applicable specifications) (Section 4.7.14 in the HTML 5.0 specification, currently a Candidate Recommendation of the W3C as of this writing [HTML5:spec]).

The adoption of a new standard does take years—on average about a decade and a half according to Andrew Odlyzko. MathML is thus not doing so badly. The example of TEX which may be thought to have taken the mathematical community by storm can be said to have taken over a decade to really catch on from its first edition in SAIL in 1978, or from the first complete rewrite in Web, a literate programming extension of Pascal. TEX did have the advantage of being a complete package for document composition not a specification attempting to be part of a larger context of specifications and technology still under very active development. It was also the work of a single genius, Donald Knuth, and not of a changing committee.

However, MathML is being utilized, not just by MR and ZB, who have been relatively early adopters as they were for TeX but by publishers with XML workflows, and by those who have to have the assurance of using a publicly adopted standard, not a proprietary one, which is coherent with the Web. While rendering support within browsers remains a problem area despite years of lobbying, it is improving—mathematics, and indeed technical documentation, is just a much less lucrative business than advertising and entertainment. In the meantime a technology originally intended as a stopgap, namely MathJax [MathJax:website], has brilliantly provided, through JavaScript, rendering support of a surprisingly high quality and reliability in almost all browsers. Perhaps its initial success can be ascribed to its being a single person's conception and work: Davide Cervone, then advised by his friend Robert Miner, who contributed a great deal to the development of MathML and regrettably died young.

4 Networks of Mathematics

The third example of software in the service of mathematical knowledge is the use of programs to analyze the nature of our subject as represented by its literature. Possibly the oldest consideration of this sort is the Erdős number, which comes from the co-authorship graph of mathematical papers. Later and more thorough analyses have been done of other networks representing mathematics' publications, whether in terms of co-authorship or co-citation, or in relation to subject areas (using the MSC), e.g. [Brunson:2013]. Further studies have begun [Dubois:2013] [Borjas:2012], leading to such modern topics as persistent homology [Bampasidou:2014] and A-theory [Babson:2006][Atkin:1974]. Machine processing of the corpus of mathematics as a natural language has also started. Analysis of the use of formulas depends on a standard notation such as MathML.

5 Conclusion

Finally let it be pointed out that the MSC and MathML are already extensively used in such places as [Wikipedia], [PlanetMath], and the [EuDML] as well as essentially in the publishing world, MathSciNet and zbMATH. The easing of access to recorded mathematical knowledge offered a possible World Heritage DML, and even use of MSC and MathML in [swMATH], make clear that mathematical knowledge management, even in its primitive present form, can aid further development of mathematics. The examples above are just starting points.

References

[Atkin:1974] Atkin, R.H.: An algebra for patterns on a complex, I. Internat. J. Man-Machine Stud. 6, 285–307 (1974); II. Internat. J. Man-Machine Stud. 8, 483–448 (1976)

[Babson:2006] Babson, E., Barcelo, H., de Longueville, M., Laubenbacher, R.: Homotopy theory of graphs. J. Algebr. Comb. 24, 31–44 (2006), doi:10.1007/s10801-006-9100-0

[Bampasidou:2014] Bampasidou, M., Gentimis, T.: Modeling collaborations with persistent homology. Preprint arXiv:1403.5346v1

[Borjas:2012] Borjas, G., Doran, K.: The Collapse of the Soviet Union and the Productivity of American Mathematicians. Quarterly Journal of Economics 127(3), 1143–1203 (2012)

[Brunson:2013] J.C. Brunson, S. Fassino, A. McInnes, M. Narayan, B. Richardson, C. Franck, P. Ion and R. Laubenbacher: Laubenbacher: Scientometrics (2013), doi: 10.1007/s11192-013-1209-z

[Dubois:2013] Dubois, P., Rochet, J.-C., Schlenker, J.-M.: Productivity and Mobility in Academic Research: Evidence from Mathematicians. IDEI Working Paper, n. 606 (October 2010), (revised March 2013)

[HTML5:spec] Berjon, R., Faulkner, S., Leithead, T., Navara, E.D., O'Connor, E., Pfeiffer, S., Hickson, I.: HTML 5, A vocabulary and associated APIs for HTML and XHTML. W3C Candidate Recommendation (February 04, 2014), `http://www.w3.org/TR/html5/`

[MathML3:spec] Carlisle, D., Ion, P., Miner, R.: Mathematical Markup Language (MathML) Version 3.0, 2nd edn. W3C Recommendation (April 10, 2014), `http://www.w3.org/TR/2014/REC-MathML3-20140410/`

[Peano:1894] Peano, G.: Formulaire de mathématiques. t. I-V. Turin, Bocca frères, Ch. Clausen, (1858-1932) 1894–1908

[Schröder:1897] Schröder, E.: Über Pasigraphie, ihren gegenwärtigen Stand und die pasigraphische Bewegung in Italien. 147–162 of Verhandlungen des ersten Internationalen Mathematiker-Kongresses in Zürich vom 9. bis 11. August (1897)

[XML:spec] Bray, T., Paoli, J., Sperberg-McQueen, C.M., Maler, E., Yergeau, F. (eds.): Extensible Markup Language (XML) 1.0 , 5th edn. W3C Recommendation, (November 26, 2008), `http://www.w3.org/TR/xml/`

[**Web References**] valid 15 May 2014 or thereafter

[EuDML] EuDML - The European Digital Mathematics Library, `https://eudml.org/`

[MathJax:website] MathJax Home Page, `http://www.mathjax.org`

[MathSciNet:website] Database, `http://www.ams.org/mathscinet/`

[MSC2010:website] Mathematics Subject Classification, `http://msc2010.org`

[MSC2010:MediaWiki] `http://msc2010.org/mscwiki/`

[MSC2010:SKOS] `http://msc2010.org/resources/MSC/2010/info/`

[MSC2010:TiddlyWiki] `http://msc2010.org/MSC-2010-server.html`

[PlanetMath] PlanetMath: math for the people, `http://planetmath.org/`

[SKOS:website] `http://www.w3.org/2004/02/skos/`

[swMATH] An information system for mathematical software, `http://swmath.org/`

[W3C:website] World Wide Web Consortium, `http://www.w3.org/`

[Wikipedia] Wikipedia, `http://wikipedia.org/`

[zbMATH:website] Database, `http://www.ams.org/mathscinet/`