

L^AT_EX to XML/MathML for Web-Based Publication*

Frederick H. Bartlett
Springer-Verlag New York, Inc.
fredb@springer-ny.com

Abstract

At Springer-Verlag, we have been frustrated for some years now with the difficulty of putting mathematics into a web-friendly format. We have not yet found a magic bullet, but . . .

The XML application MathML may be the first real tool for putting mathematics on the Web in a useful form. Suppliers of mathematical tools such as Mathematica and Maple are gearing up to use MathML as an input/output format; thus, we can look forward to a day when mathematics on the Web will be truly interactive.

It is likely that — even if MathML fulfills every bit of its promise — T_EX will continue to be used for the preparation of mathematics for display and printing.

This presentation is an account of our efforts to translate author-generated L^AT_EX into XML. The project can be divided into four stages:

1. Normalizing (La)T_EX. That is, transforming authors' idiosyncratic usages (and even more idiosyncratic macro definitions) into consistent, and consistently structured, files. The vast majority of author-generated L^AT_EX files can be converted easily with a minimal understanding of TeX's digestive tract; those which can't (especially plain T_EX files) will require some human intervention — or increasingly sophisticated (read 'bloated') software.
2. Converting to XML. This is the easy part: changing structural L^AT_EX tags into XML tags.
3. Converting to MathML. And this is the hard part: It would be ideal to be able to convert L^AT_EX math into both presentation and content MathML coding. Unfortunately, this is, even in principle, extremely difficult. So at first we concentrate on the L^AT_EX-to-presentation mark-up path. Eventually, it will be possible to produce an interactive L^AT_EX-to-content mark-up converter for authors.

4. Going backwards. It will eventually be helpful to authors and publishers if MathML/XML can be converted back to (La)T_EX, but this is not a high priority at the moment.

This talk describes something that is very much a work in progress, so a discussion period will be most welcome.

* [No Paper Submitted. – Ed.]