Very Like a Nail: Typesetting SGML with TEX*

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 — TEX 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 LATEX into XML. The project can be divided into four stages:

1. Normalizing (La)TEX. That is, transforming authors' idiosyncratic usages (and even more idiosyncratic macro definitions) into consistent, and consistently structured, files. The vast majority of author-generated LATEX files can be converted easily with a minimal understanding of TeX's digestive tract; those which can't (especially plain TEX files) will require some human intervention — or increasingly sophisticated (read 'bloated') software.

2. Converting to XML. This is the easy part: changing structural LATEX tags into XML tags.

3. Converting to MathML. And this is the hard part: It would be ideal to be able to convert LATEX math into both presentation and content MathML coding. Unfortunately, this is, even in principle, extremely difficult. So at first we concentrate on the LATEX-to-presentation mark-up path. Eventually, it will be possible to produce an interactive LATEX-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)TEX, 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.

To TEX or not to TEX

To TEX, or not to TEX: that is the question:
Whether 'tis nobler on the page to suffer
The slings and arrows of outrageous software,
Or to write code against a sea of troubles,
And by opposing end them? Use Word? Use Quark?
No more! For such as they could never end
The heartache and the thousand unnatural shocks
That type is heir to.
    LATEXe*?
Devoutly to be wish'd! Or Quark to TEX?
To TEX? Now there's a dream. And here's the rub:
From that disguisèd TEX the dream may come
That TEX should shuffle off this mortal coil,
So should we pause? There's no respect
For TEX in all of its long life;
For who would bear the whips and scorns of Frame,
WordPerfect's wrong, Microsoft's contumely,
The pangs of despis'd TEX, Incontext's delay,
The insolence of Active TEX and the spurns
That patient Wizards of th' WYSIWYGers take,
When he himself might his quietus make
In a plain TEX style? Who would authors bear,
To grunt and sweat under a weary life,
But that the dread of something after TEX,
The undiscover'd standard from ISO
And W3C, puzzles the will
And makes us rather love the type we have
Than fly to others that we know not of?
    TEX's enterprise of great pith and moment
With XML its currents join anon,
And gain the funds of moguls.

—stolen by Fred Bartlett

---

* [No paper submitted. – Ed.]