Toward a Mathematical Markup Language

Ralph Youngen

The HTML-Math Working Group of the World Wide Web Consortium (W3C) recently released a Working Draft of a Mathematical Markup Language (MathML) for delivering mathematical content in Web documents. The full text of this Working Draft is available at `http://www.w3.org/TR/WD-math/`. What follows is the Executive Summary, which is meant to serve as an introduction to the proposal outlined in the full Working Draft and to provide a framework to assist the reader in understanding the history, current state, and future of the MathML specification.

MathML Background

Efforts to define a specification for mathematics in HTML have been under way in the W3C for several years. Dave Raggett made a proposal for HTML-Math in 1994. Although that proposal was not ultimately incorporated into HTML 3.2, a panel discussion at the WWW IV Conference in April 1995 demonstrated that there was clear interest in pursuing a specification for mathematics on the Web. A group was formed to discuss the problem further.

In the intervening two years the W3C and the math working group have both evolved substantially. The small informal math group has grown and has been formally reconstituted as the W3C HTML-Math Working Group. The current Working Group is composed of experts from commercial publishers and software vendors, as well as not-for-profit publishers and research organizations hailing from both Europe and North America.

Early in the Working Group's discussions a few fundamental observations were noted:

- The potential audience for a mathematics specification for the Web is quite large, encompassing students and teachers at nearly all educational levels, professional researchers from many different scientific disciplines, software vendors, scientific publishers, etc.
- Contemporary technologies for rendering mathematics on the Web (e.g., as in-line graphics or using complex HTML font and positional commands) contain drawbacks serious enough to prevent their widespread use.

As the Working Group began its progress toward a more robust method of rendering mathematics in the Web environment, it also became evident that the motivation within the Group for defining a mathematics specification for the Web was emanating primarily from two distinct sources:

- From publishers who have an immediate need for displaying mathematical content on the Web. Publishers have legacy data in TEX and/or SGML and presently have very limited options for displaying such documents on the Web.
- From software vendors who wish to link their products for manipulating mathematical expressions with mathematics in Web pages. This might be accomplished by cutting and pasting a mathematical expression from a Web page into the mathematical software or by other, more direct means.

The MathML specification directly responds to both of these needs. MathML provides both a Presentation Tag set to specify how mathematics should be rendered within a Web page and a Con-

Ralph Youngen is the director of Electronic Product Development for the AMS and a member of the HTML-Math Working Group of the World Wide Web Consortium. His e-mail address is `rey@ams.org`.

Example of MathML Encoding for $x^2 + 4x + 4 = 0$

Presentation Tags

```
<MATH>
<MROW>
    <MROW>
        <MSUP>
          <MI>x</MI>
          <MN>2</MN>
        </MSUP>
        <MO>+</MO>
        <MROW>
          <MN>4</MN>
          <MO>&InvisibleTimes;</MO>
          <MI>x</MI>
        </MROW>
        <MO>+</MO>
        <MN>4</MN>
    </MROW>
    <MO>=</MO>
    <MN>0</MN>
</MROW>
</MATH>
```

Content Tags

```
<MATH>
<EXPR>
    <EXPR>
        <EXPR>
          <MI>x</MI>
          <POWER/>
          <MN>2</MN>
        </EXPR>
        <PLUS/>
        <EXPR>
          <MN>4</MN>
          <TIMES/>
          <MI>x</MI>
        </EXPR>
        <PLUS/>
        <MN>4</MN>
    </EXPR>
    <E/>
    <MN>0</MN>
</EXPR>
</MATH>
```

tent Tag set that can be used to capture the meaning of the mathematics being encoded. These tag sets are further explained below.

The MathML Specification

MathML itself is a combination of the Presentation Tag and Content Tag sets. MathML is an XML application that adheres to the current XML specification, which itself is a Working Draft under the W3C. (XML is a proposed extension to HTML for delivering content on the Web. XML is designed to interoperate with current HTML specifications. See `http://www.w3.org/TR/WD-xml.html` for the full XML Working Draft.)

The MathML Presentation Tags correspond to the constructors of traditional math notation while also providing mechanisms to fine-tune the visual rendering of mathematics in a Web environment. MathML Presentation elements are at least as powerful as TEX in their ability to control the visual representation of math. MathML Presentation Tags can therefore be used to render sophisticated research-level mathematical notation on the Web.

The MathML Content Tags are intended to support the encoding of the underlying mathematical content of an expression. Mathematics encoded with Content Tags could, for example, be cut from a Web document and pasted into mathematical software (such as computer algebra systems) where the mathematics could be further manipulated. MathML Content Tags should be adequate for coding of most formulas used in education through the U.S. high school level and probably beyond that, through the first two years of college.

MathML Presentation Tags and Content Tags may be mixed within a single mathematical expression. By proper use of Presentation Tags alone, the mathematical structure of an expression can be represented well enough to permit its manipulation by computer algebra systems, though such systems must make assumptions about the meanings of individual symbols in some instances. When a potential for ambiguity exists, Content Tags may be employed to precisely specify the intended mathematical meaning. An author may therefore choose any point along this spectrum from Presentation to Content Tags according to how precisely (s)he wants to specify the mathematical content of an expression.

Whether using Presentation Tags or Content Tags, MathML is designed to be mixed in with ordinary HTML commands, just as markup for frames mixes with HTML, for example. However, MathML markup is verbose and is not optimized for hand entry. While some simple examples may be keyed by hand, it is likely that most MathML markup will be generated either by translators/filters that convert from other mathematical markup languages (e.g., TEX, some SGML specifications, etc.) into MathML or by authoring tools that output MathML

directly. Translators from SGML or TEX would probably choose to convert these formats into MathML Presentation Tags, while WYSIWYG authoring tools may choose to output either MathML Content Tags or MathML Presentation Tags.

The HTML-Math Working Group fully recognizes the need for simplified mathematical input syntaxes that are better suited for human entry in addition to the machine-friendly MathML notation. The Working Group plans to focus on such input syntaxes after the completion of the MathML specification.

Embedding MathML in HTML Pages

Within an HTML document, MathML notation appears between the tags $...$ (for in-line math) and <MATHDISP>...</MATHDISP> (for display math). The box on the preceding page illustrates MathML encoding using both Presentation Tags and Content Tags for the expression $x^2 + 4x + 4 = 0$.

The top-level HTML tags <F>...</F> and <FD>...</FD> are also being reserved for alternate mathematical notations (i.e., notations other than MathML). <F>...</F> and <FD>...</FD> are mnemonic for "formula" and "formula display", analogous to the <MATH> and <MATHDISP> tags. For example, a plug-in capable of rendering TEX code directly might be invoked as follows:

```
< FD TYPE="application/x-tex">
  2x^2+3y=z</FD>
```

Initially plans are under way to implement MathML renderers as embedded elements such as plug-ins and Java applets. Both IBM's techexplorer (a plug-in), at `http://www.ics.raleigh.ibm.com/ics/techexp.htm`, and WebEQ (a Java applet) from the Geometry Center at the University of Minnesota, at `http://www.geom.umn.edu/software/WebEQ`, have stated that they plan to support MathML in upcoming versions. Both are freely available.

MathML Timetable

The HTML-Math Working Group issued the initial MathML Working Draft on May 15, 1997. Several revisions are expected through the summer and fall of 1997. The Working Group has a goal of promoting the Working Draft to a W3C Proposed Recommendation around October 1997.

After the MathML specification is ratified by the W3C, the HTML-Math Working Group then plans to turn its attention to several other matters, such as simplified input syntaxes suitable for hand-editing elementary mathematical notation in the <F> and <FD> tags and macro mechanisms for MathML itself. A second Working Draft covering these and other considerations is planned for May 1998.