PUBLISHING MATHEMATICS ON THE WEB
A Survey of Simple, Inexpensive Methods

Harja Santana Purba[1] & Abdullah Embong[2]

[1,2] *School of Computer Science, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia*
[1] harja63@yahoo.com; [2]ae@cs.usm.my

Abstract

There are a variety of approaches and technologies for publishing mathematics document on the Internet. In this paper we highlight three different methods, which have different purposes and applications. They are: direct embedding in HTML, page-oriented format, and advanced markup languages. The first method is suitable for integrating mathematical expression with other media such as movies, sounds, interactivity, data access, etc. The second method is used when we want post mathematical material in a form that is as close as possible to the original format, such as for printing or display locally. The last method is especially useful for integrating mathematical material with other Web technology, such as scripting languages. There are many tools available for authoring mathematical expression with little or no cost. The main tools are needed for encoding and validating HTML, CSS, XML, and MathML documents. Authoring mathematical document also needs an equation editor for creating mathematical expression. Other tools are needed for editing and converting mathematical document to eBooks format, such as PDF, LIT, HLP, and CHM generator.

Keywords: HTML, CSS, XSL, XML, XHTML, MathML, PDF, HLP, CHM, equation editor, MathML editor.

1. Introduction

The Internet is an invaluable tool for communicating scholarly information, but there remain difficulties and issues when trying to use the Web to communicate certain kinds of information, most notably scholarly papers and journal articles containing complex mathematical symbols and formulae. Currently there is no single, universally accepted technology that could be used. What we need is a comprehensive digital mathematical library, which can accommodate a variety of technologies and approaches. There is no single right answer for all of the contexts. This paper will study three different options, which are available and highlight the strengths and weaknesses of each. The three options for putting mathematics on the web pages are: Direct Embedding in HTML, Page-Oriented Format, and Advanced Markup Languages.

2. Direct Embedding in HTML

HTML [1] is especially ubiquitous and is also very much an open standard. As such it has advantages of economy and essentially universally available to end-users. In native HTML many simple, one-line mathematical expressions can be inserted into a web page by appropriate use of italics, horizontal spacing, superscripts and subscripts, all of which are a standard in HTML, and therefore my be viewed with any browser. HTML provides intrinsic encoding for a limited number of special characters, tags to designate pre-defined super and subscript character positioning, and a tag to turn off automatic line breaks, and little else. However, direct support for mathematics in HTML is minimal. There's no direct HTML support for kerning, the use of combining diacritics, arbitrary glyphs and fonts, equation structures, or any of the other rendering system features essential for presenting complex mathematics.

The advent of Cascading Style Sheets and font embedding has extended the ability to depict mathematics in HTML, but such techniques are limited. CSS [2] was not designed with mathematics in mind and CSS positioning mechanics are clumsy and awkward to apply for mathematics. Robust generic solutions are impossible and the process is generally quite labor intensive. Moreover, both CSS and embedded font solutions vary by Web browser type, and often by Web browser version.

The other approach used in HTML is to embed mathematics needed for a Web page in the form of GIF, JPEG, or PNG images. This approach, though ubiquitous also has severe limitations. Reducing mathematics to a binary screen display format means it is of no use for resource discovery purposes. Embedding images in the midst of sentences can complicate document full-text search and discovery generally. The number of images required for even a modest paper can be quite high. Binary display representations of mathematics don't scale as screen fonts are changed, and typically don't match the fonts used on different clients well.

Putting mathematics using direct embedding in HTML has several advantages. Because the format is HTML-based, math can more easily be combined with other media such as movies, sounds, interactivity, data access, etc. This method doesn't require server-side support. There is no need for fonts to be installed, which is important for cross-platform browsing, but this approach has several disadvantages. Interaction with math is limited to scripts selection among pre-composed equation images. Another limitations is ugly for printing because GIFs are low-res screen images

There are many good HTML authoring tools available at little or no cost. Table 1 bellows is a summary of tools, which available freely.

Table 1: HTML Authoring Tools

	Authoring Tools
HTML Editor	HTML-Kit
HTML Validator	HTML Tidy (HTML-Kit Plugin)
CSS Editor	TopStyle
Equation Editor	MathCast

HTML-Kit [3] is a powerful HTML editor with plug-in library for validation HTML, XML and XHTML tags. HTML-Kit also has capability for editing CSS file with several limitations. For readability purposes CSS file can be created with TopStyle [4]. TopStyle CSS editor have lite version which available freely. MathCast [5] is an equation editor, an application that allows us to input mathematical equations. These equations can be used in written documents, web pages, and even databases. They could be rendered graphically to the screen, to picture files (.bmp, .png, and .emf files), or to MathML [6] - today's leading standard language for describing mathematics. MathCast is a free and open source application. This software is licensed under GPL.

3. Page Oriented Format

Page oriented format is used when we want to post mathematical material in a form that is as close as possible to the original. In this situation the e-Books format is exactly what we need. There are many types of e-Books format which is available, such as PDF, LIT, HLP and CHM. PDF is an open standard, and anyone may write applications that can read or write PDF with royalty-free. LIT is an e-Books format developed by Microsoft. HLP is a standard for previous Microsoft Windows Help file. The current Windows Help standard is HTML Help that compressed HTML files in CHM file format, which is organized in a database-like manner. It has set a set of web pages written in a subset of HTML and a hyperlinked table of contents.

The challenges are associated e-Books with mathematical typography. Of course the challenges are nothing new. Mathematicians have been thinking of print technologies since printing was invented. As a solution to the problem of mathematics on the printed page, TEX [7] has proven a popular and relatively satisfactory solution since its introduction in 1978. Use of it in a Web environment leverages the extensive work that has been done in print publishing and typesetting environment. It is extensible and there are a wide range of TEX-aware authoring and publishing tools. Really more a set of instructions for presentation than a self-contained binary version of a mathematical display, TEX is most often implemented on the Web only after translation to DVI and then another Web ubiquitous format such as Adobe PDF [8].

TeX system has its drawbacks and limitations. It is complicated to learn, somewhat difficult for new user to use well, and it comes in a multiplicity of flavors, confusing to the non-specialist. It was developed and optimized for the printed page. TEX and PDF are particularly good for mimicking the format and layout of a printed journal article, but as we move to more and more "born-digital" content this printed page orientation has drawbacks. Also, semantic content is limited. In its native form, some semantics which are implicit in an equation is encoded in TEX, but it was not designed to describe the meaning of the mathematics, only the way the mathematics should look when printed. Once transformed to DVI and then to PDF, even more of the semantic meaning is lost.

MiKTeX [9] is an up-to-date implementation of TeX and related programs for Windows on x86 systems. WinEdt [10] provides an integrated development environment and is widely used as a front-end for compilers and typesetting systems, such as MiKTeX. Using

MiKTeX with WinEdt editor shell has a strong predisposition towards the creation of TeX documents. WinEdt can connect to Adobe Reader [11] for displaying PDF file directly. WinEdt's highlighting schemes can be customized for different modes. WinEdt's Spell Checking functionality supports multi-lingual setups, and dictionaries (word-lists) are available for many languages.

Recently the most exciting innovation is for making PDF file from popular word processor, e.g. Microsoft Word [12]. MathCast equation editor can be used for creating mathematical expression in image format, which can be embedded with Word document. There are several printer utilities to convert Word document directly to PDF file, e.g. PDFCreator [13]. PDFCreator is an open source application that can create PDF's from just about any program that prints using windows printers. PDFCreator utilize Ghostscript [14] to create PDF file. With PDFCreator we can create PDF's, Postscript and Encapsulated Postscript files, create images from our documents, combine documents and much more.

There are several advantages for making e-Books in PDF format. This format is good for online delivery of documents whose primary publishing medium is paper. PDF format is also good quality formatting and printing that is faithful to the designer's original intent. Another advantage is this format doesn't require server-side support. If math fonts are embedded in the document, math symbol display is guaranteed to work.

The approach has its drawbacks and limitations. Because the format is not HTML-based, math can't be combined easily with other media such as movies, sounds, interactivity, data access, etc. The Web browser is merely hosting the PDF viewer application, which must integrate its own user interface with that of the browser. This can be more cumbersome or confusing for the user and limits interaction. Changes made using the browser's interface - e.g. Text Smaller or Larger - have no effect on the PDF display. Page-oriented formats like PDF (TeX and PS, etc.) will continue to be especially useful when doing retrospective digitization of print materials, but we can improve on these formats for new, born-digital mathematics.

The simplest one for creating eBooks may be in LIT format. Now anyone can create eBooks from Microsoft Word files at the click of a button. The Read in Microsoft Reader [15] add-in for Microsoft Word enables us to convert any Word document into a Microsoft Reader [16] (.LIT) format eBook in just a few simple steps. The Read in Microsoft Reader add-in works with Microsoft Word 2000, Microsoft Word XP and Microsoft Word 2003. It requires Windows 98, Windows Me, Windows NT 4.0 with Service Pack 6 or higher, Windows 2000, or Windows XP in order to run [17]. All the tools including Read in Microsoft Reader, Microsoft Reader and MathCast, but Microsoft Word are freely available.

Another kind of eBooks can be made with WinHelp (.HLP) format. WinHelp refers to Help files developed with Rich Text Format (RTF) files and Microsoft Windows Help Compiler (HHW). Since 1995, Help developers could choose between WinHelp 3 for Windows 3.x system and WinHelp 4 for Windows 95/98/NT. Microsoft Word has included RTF converter in their distribution, which can be used for creating RTF files. The

WinHelp authoring tool, Microsoft Help Workshop [18], provides an easy-to-use system for creating and managing help projects and their related files.

HTML Help (.CHM) is the current Windows Help standard, combining the functionality of WinHelp with the flexibility of Hypertext Markup Language (HTML) and the power of ActiveX controls and script. Microsoft HTML Help consists of an online Help Viewer, related help components, and help authoring tools from Microsoft Corporation. The Help Viewer uses the underlying components of Microsoft Internet Explorer to display help content. It supports HTML, ActiveX, Java, scripting languages (JScript, and Microsoft Visual Basic Scripting Edition), and HTML image formats (.jpeg, .gif, and .png files). The help-authoring tool, HTML Help Workshop [19], provides an easy-to-use system for creating and managing help projects and their related files.

There are many authoring tools for creating e-Books which are available freely. Table 2 is a summary of authoring tools which are available freely, except for Microsoft Word and WinEdt. WinEdt is distributed as shareware that can be registered with little cost.

Table 2: e-Books Authoring Tools

e-Books Format	Authoring Tools	e-Books Reader
PDF	MikTeX, WinEdt, Ghostscript	Adobe Reader
	MsWord, Equation Editor, PDFCreator, Ghostscript	
LIT	Microsoft Word, Read in Microsoft Reader, Equation Editor	Microsoft Reader
HLP	Microsoft Help Workshop, MsWord RTF Editor, Equation Editor	None, but Internet Explorer (version 4 or later) must be installed on the system.
CHM	HTML Help Workshop, HTML Editor, HTML Validator, CSS Editor, Equation Editor	None, but Internet Explorer (version 4 or later) must be installed on the system.

4. Advanced Markup Languages

The other option is Advanced Markup Languages which is implemented from XML. XML has found much wider acceptance among Web users. A new generation of specialized XML implementations has appeared, including MathML. The problem is not all web browsers support its format directly. Of the major web browsers, those that directly support the format are recent versions of Mozilla [20] and its derivatives. Other browsers support the format if external plug-in are installed. For example, Internet Explorer supports MathML with a MathPlayer [21] plug-in. Our aim is to find a universally accepted format which is cheap to implement and is supported by all of the major web browsers.

Authoring mathematics document in HTML with MathML enable has several advantages. Since the page is HTML-based, it is easy to integrate with other Web technologies.

Interaction is snappy as it occurs on the client and doesn't require server-side support. Interaction is powerful as scripts can generate math to be displayed on the fly. The approach has its drawbacks and limitations. Interface between math rendering component and page requires that the width and height of each equation be known in advance. This is difficult to do if the math is generated by scripting. Math pages containing hundreds of individual equations can cause performance problems due to the inability of browsers to handle that many component instances. This issue can be minimized by using components in conjunction with GIFs for non-interactive equations.

In Netscape 7.2 [22], Firefox 1.0.1 [23] and Mozilla 1.7.5, MathML support is built into the rendering engine. In term of speed and quality, it is comparable to the rest of the browser text. Since it is built in, users do not need to download a separate plug-in. However, this browser found that they need to download and install math fonts. MIT MathML fonts [24] can be downloaded freely. The MIT MathML Font package includes the following fonts: Mathematica 4.1 TrueType, TeX Computer Modern TrueType, MT Extra. In Internet Explorer, MathPlayer provides MathML support. MathPlayer utilizes powerful, low-level extension capabilities called *behaviors* only available in the Windows version of Internet Explorer [25]. However, by utilizing behaviors, MathPlayer achieves high performance, native-quality rendering and seamless browser integration. MathPlayer is installed by downloading a standard Windows installer. The installer also includes the fonts needed by MathPlayer.

There are many tools for authoring web pages with MathML support which are available at no cost. Table 3 shows a summary of authoring tools which are available freely.

Table 3: HTML+MathML Authoring Tools

	Authoring Tools
HTML, XML, XHTML Editor	HTML-Kit
HTML, XML, XHTML Validator	HTML Tidy (HTML-Kit Plugin)
CSS Editor	TopStyle
MathML Equation Editor	MathCast

The main technologies for web pages with MathML enable are HTML Platform [26]. The HTML Platform are XML [27], HTML, and MathML for encoding content, XSL [28] and CSS for styling and processing documents, and JavaScript and DOM for scripting of dynamic features in a page. The browser support for HTML Platform varies. Table 4 summarizes browser support for this platform.

Table 4: HTML+MathML Browser Support

Browser	**MathML Support**	**Math Font Support**
Internet Explorer 6.0+	MathPlayer Plugin	Built-in with MathPlayer
Netscape 7.1+	Built-in with Browser	MIT MathML Font
Mozilla 1.1+	Built-in with Browser	MIT MathML Font
Firefox 1.0+	Built-in with Browser	MIT MathML Font

In order to maximize the number of platforms it will be viewable on, the W3C Math Working Group [29] recommend some rules for writing HTML document with MathML

enables. Firstly, the web page should be written using XHTML with the MathML markup inlined, as in the following example:

```
<?xml version="1.0"?>
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>...</head>
  <body>
    <h1>Example</h1>
    ....
    <math xmlns="http://www.w3.org/1998/Math/MathML">
      <mi>x</mi><mo>+</mo><mn>3</mn>
    </math>
  </body>
</html>
```

Secondly, the web document should be added by a stylesheet processing instruction. The following line, in bold characters, should be inserted at the beginning of the XHTML [30] page, before the opening <html> tag but after the XML declaration <?xml...?> if it exists:

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl"
href="http://www.w3.org/Math/XSL/mathml.xsl"?>
<html xmlns="http://www.w3.org/1999/xhtml">
...
```

If the document containing math is meant to be viewed locally, without an internet connection, the declaration should be

```
<?xml-stylesheet type="text/xsl" href="mathml.xsl"?>
```

And a copy of the following files should reside on the local computer, along with the main document: **mathml.xsl**, **pmathml.xsl**, **ctop.xsl** and **pmathmlcss.xsl** [31]. If the document contains only Presentation MathML [32], only the files **pmathml.xsl** and **pmathmlcss.xsl** are needed. The display will be faster if only those two files are used.

5. Conclusions

There are variety of approaches and technologies for publishing mathematics document on the Internet. The three options are: direct embedding in HTML, page-oriented format, and advanced markup languages. The first method is suitable for integrating mathematical expression with other media such as movies, sounds, interactivity, data access, etc. The second method is used when we want post mathematical material in a form that is as close as possible to the original format, such as for printing or display locally. The last method is especially useful for integrating mathematical material with other Web technology, such as scripting languages.

References

1. HTML, http://www.w3.org/MarkUp/, accessed on April 15, 2005
2. CSS, http://www.w3.org/Style/CSS/, accessed on April 15, 2005
3. HTML-Kit HTML Editor, http://htmlkit.com, accessed on March 24, 2005
4. TopStyle CSS Editor, http://www.Bradsoft.com, accessed on March 24, 2005
5. MathCast Equation Editor, http://mathcast.sourceforge.net/home.html, accessed on April 22, 2005
6. MathML, http://www.w3.org/Math/, accessed on April 22, 2005
7. TeX (TUG: TeX Users Group), http://www.tug.org, accessed on March 17, 2005
8. PDF, http://www.adobe.com/acrofamily/main.html, accessed on March 17, 2005
9. MikTeX, http://www.miktex.org, accessed on March 17, 2005
10. WinEdt, http://www.winedt.com, accessed on March 17, 2005
11. Adobe Reader, http://www.adobe.com/product/acrobat/readermain.html, accessed on March 24, 2005
12. Microsoft Word, http://www.microsoft.com, accessed on March 10, 2005
13. PDFCreator, http://sourceforge.net/projects/pdfcreator/, accessed on April 10, 2005
14. Ghostscript, http://www.cs.wisc.edu/~ghost/, accessed on March 17, 2005
15. Read in Microsoft Reader, http://www.microsoft.com/downloads/, accessed on March 10, 2005
16. Microsoft Reader, http://www.microsoft.com/downloads/, accessed on March 10, 2005
17. Windows (98, Me, NT 4.0, 2000 and XP), http://www.microsoft.com, accessed on March 10, 2005
18. Microsoft Help Workshop, http://www.microsoft.com/downloads/, accessed on February 23, 2005
19. HTML Help Workshop, http://www.microsoft.com/downloads/, accessed on May 10, 2005
20. Mozilla, http://www.mozilla.org, accessed on February 10, 2005
21. MathPlayer, http://www.dessci.com/en/products/mathplayer/, accessed on April 22, 2005
22. Netscape, http://channels.netscape.com/ns/browsers, accessed on February 10, 2005
23. Firefox, http://www.mozilla.org, accessed on February 10, 2005
24. MIT MathML Fonts, http://web.mit.edu/is/topics/webpublishing/mathml/fonts-win.html, accessed on April 15, 2005
25. Microsoft Internet Explorer, http://www.microsoft.com/windows/IE/, accessed on February 10, 2005
26. R. Miner and P. Topping, *Math on the Web: A Status Report.* (Design Science, Long Beach, January, 2001). url: http://www.dessci.com/webmath/status
27. XML, http://www.w3.org/XML/, accessed on April 15, 2005
28. XSL, http://www.w3.org/Style/XSL, accessed on April 15, 2005
29. The W3C Math Working Group, http://www.w3.org/Math/, accessed on April 10, 2005
30. XHTML, http://www.w3.org/MarkUp/Overview.html, accessed on April 22, 2005
31. The W3C XSL Files, http://www.w3.org/Math/XSL/, accessed on April 22, 2005
32. Presentation MathML, http://www.w3.org/Math/TR/MathML2/chapter3.html, accessed on April 22, 2005