

htexml: A Web Publishing Tool for STEM Authors

Cesar O. Aguilar

Associate Professor

Department of Mathematics, SUNY Geneseo

Abstract

The problem that this proposal seeks to solve is to provide STEM authors with an easy to use and freely available web-based application that converts LaTeX documents into modern and mobile-ready HTML web pages. The proposed software will allow instructors and researchers to quickly and easily publish and share open-access STEM textbooks, course notes, and research articles on their personal websites or on a dedicated website similar to arxiv.org. In contrast to the limitations imposed by publishing documents on the web as PDF files, the software tool will leverage modern web technologies to create websites that are responsive to varying screen sizes, provide superior inter-document navigation, and improve the usability of assistive technologies such as screen readers and Braille displays.

1. PROJECT DESCRIPTION

Recent web technologies have made it easier for STEM authors to publish HTML documents on the web containing rich mathematical content. Specifically, the JavaScript library **MathJax**¹ allows authors to include LaTeX code directly in HTML source files which is then parsed and rendered natively on a web browser. In addition, the mathematical content produced by MathJax includes accessibility features such as speech and Braille support and sub-expression highlighting and magnification.

How does one use MathJax? The simplest way is to include a script tag in the head of an HTML document that fetches the MathJax library and thereby allowing the document creator to mix HTML markup with LaTeX markup in a single HTML file; the HTML markup is handled by the browser on page load and the LaTeX markup is processed by MathJax to dynamically produce mathematical expressions.

The main barrier to using MathJax as just described above is that STEM authors are trained in, and prefer to use, LaTeX to create mathematically rich documents rather than HTML. The usual end product of a LaTeX document is a PDF file which is subsequently shared on the web. However, there are two principal disadvantages with disseminating documents on the web using the PDF file format, and in particular, documents containing rich mathematical content:

1. **PDF files are not internet friendly.** *“PDFs are meant for distributing documents that users will print. They’re optimized for paper sizes, not browser windows or modern device viewports. We often see users get lost in PDFs because the print-oriented view provides only a small glimpse of the content. Users can’t scan and scroll around in a PDF like on a web page.”*². In short, PDF files were never meant to be used on the web to view content, especially now with the ubiquity of mobile phones and small tablets.

¹<https://www.mathjax.org>

²[Avoid PDF for On-Screen Reading](#)

2. Without additional significant effort by a writer, **screen readers cannot reliably parse the content of a PDF file that contains rich mathematical content**. This presents a serious barrier, for example, to students that have physical disabilities and require a specialized screen reader to study from textbook sources. The barrier is both physical and financial. On the other hand, the native language on the web for describing mathematical notation and structure, called MathML, is inherently screen reader friendly. The goal of MathML is to enable mathematics to be served, received, and processed on the World Wide Web, just as HTML has enabled this functionality for text. Alongside the visual rendering of the math content, a by-product of MathJax output is MathML code that describes the semantics of the visually rendered mathematical expressions.

The main purpose of the project would be to develop a web-based software tool to convert a source file written in LaTeX to a source file in HTML, and more generally, to convert a textbook written in LaTeX to one structured in HTML with navigation. I have used a preliminary version of the proposed software to convert a few of my LaTeX documents to HTML³.

2. BROADER IMPACTS

An instructor using the proposed software tool would be able to make available to their students resources that are:

- freely available,
- easily accessible,
- in compliance with the Web Content Accessibility Guidelines for making Web content more accessible for users with physical disabilities ([WCAG 2.1 \(link\)](#)),
- and in a format compatible with various screen sizes.

The proposed software tool would facilitate the creation of Open Educational Resources (OER) and thus increase the number of free textbooks available to students on the web.

The proposed project would include a plan to provide STEM educators with workshops on existing web technologies, and the proposed software tool, available for publishing math content on the web. The workshops would be held at regional and national conferences.

³<https://www.geneseo.edu/~aguilar/pages/notes>