Dual presentation with math from one source using GELLMU

William F. Hammond
Department of Mathematics & Statistics
University at Albany
Albany, New York 12222   USA
`http://www.albany.edu/~hammond/`

Abstract

Two traditional approaches for achieving simultaneous print and HTML output from a single marked-up source are relevant to the TeX community.

1. Write a LaTeX article, and use a program that translates to HTML.

2. Write an article in an author-level XML document type, and use standard XML translation methods to generate both LaTeX and HTML.

This article addresses a hybrid approach: the use of "generalized LaTeX", as implemented in the GELLMU Project, to produce dual presentation from a single LaTeX-like source. The method combines the reliability of XML document transformation with many of the conveniences traditionally available to LaTeX authors.

1 Introduction

A contemporary author writing an article for "dual presentation" has in mind both the classical printed presentation of an article and the online form of an article formatted in HTML. There are particular challenges when mathematics is involved because it is moderately difficult to produce correct mathematical markup for modern HTML documents.

While mathematics is a principal concern in this article, most of what is said here is relevant, though less critically so, for documents that do not involve mathematics.

There are two main approaches for achieving dual presentation that are relevant to the TeX community. (Texinfo, the language of the GNU documentation system, also provides a route for dual presentation of articles without mathematical markup.)

1. Write a LaTeX article, and use a program that translates to HTML.

2. Write an article in a suitable XML document type, such as DocBook, and use standard software for generating LaTeX and HTML.

Both methods present challenges to authors who have been accustomed to using LaTeX. In particular, if mathematics is involved, there are no widely deployed XML document types that support author-level mathematical markup. Since mid-2002 mathematical content in the second-generation form of HTML has been supported by the two most widely deployed web browsers, but not many articles seem to have appeared on the web in this form so far. The most likely reason is difficulty of creation.

This article addresses the use of "generalized LaTeX", as implemented in the GELLMU Project, to produce dual content from a single LaTeX-like source. (The overall system design in GELLMU is one for multiple outputs although at this time the standard implementation provides (i) printed output via standard LaTeX and (ii) HTML.)

A generalized LaTeX article under the GELLMU Project is essentially equivalent to an XML document under an author-level document type that may be called "GELLMU article". Preparing documents this way combines the reliability of XML document transformation with most of the conveniences, such as *newcommand* macro substitutions, the use of blank lines for paragraph boundaries, and cross-referencing, that are available when writing LaTeX markup.

It should be emphasized that the GELLMU Project does not provide translation of classical LaTeX to HTML or to any XML document type. Of course, one may open a classical LaTeX document in an editor and invest time and energy to "port" it to GELLMU source, but there is no automation for this.

The task of translating legacy documents to HTML and to XML document types is difficult. In 2007 we are witness to more than 10 years of effort in this direction, and we still have no easy path, particularly when mathematics is involved. Nonetheless, translation, to the extent that it is reasonably pos-

William F. Hammond

sible, is very important because we have 30 years of legacy documents. (Most of the past translation efforts have been aimed at standard structured formats like HTML and the *DocBook* XML document type. It would be interesting to see if the translation task toward such standard formats can be improved by first translating toward an author-level document type, such as GELLMU *article*, that models structured LaTeX rather closely and then translating from there toward the original target.)

When a contemporary author has dual presentation in mind for new documents, writing classical LaTeX is no longer the best way to proceed because of the difficulty of translation from LaTeX to other formats. Writing for a suitable author-level XML document type is a much better way to have seamless dual presentation, and using the LaTeX-like front end offered by GELLMU makes it seem to the author very much like writing LaTeX.

In his talk preceding mine at TUG 2007 Chris Rowley ventured the idea that "peak TeX", like "peak oil", lies in the near future. In saying that I hope he is speaking of a peak in relative use by authors as markup source of the print-focused typesetting language that we have known. The other side of this is that *if*, as a community, we come to appreciate the usefulness of LaTeX-like markup, as opposed to classical LaTeX, as a general author front end for writing structured documents and then come to understand that practice as part of the TeX world and hone our techniques of formatting these structured documents for classical TeX-based typesetting, the future I imagine is one of more, not less, TeX.

2 Writing source markup

This is GELLMU source for a short paragraph with a relatively simple mathematical display.

```
The following identity may be
regarded as a formulation of the
Weierstrass product for the Gamma
function.
\[ \int_{0}^{\infty} t^x
   e^{-t} \frac{dt}{t} \int:
   = \frac{1}{x}
   \prod_{k=1}^{\infty}
   \frac{\bal{1 + \frac{1}{k}}^x
   }{\bal{1 + \frac{x}{k}}}
   \prod: \]
Understanding the derivation of
this identity is reasonable for a
bright student of first year
undergraduate calculus in the
United States.
```

This source compiles to:

The following identity may be regarded as a formulation of the Weierstrass product for the Gamma function.

$$\int_0^\infty t^x e^{-t} \frac{dt}{t} \;=\; \frac{1}{x} \prod_{k=1}^\infty \frac{\left(1 + \frac{1}{k}\right)^x}{\left(1 + \frac{x}{k}\right)}$$

Understanding the derivation of this identity is reasonable for a bright student of first year undergraduate calculus in the United States.

The markup looks like classical LaTeX. In fact, except for the use of the zone closers `\int:` and `\prod:`, it would be classical LaTeX. It is *generalized LaTeX*.

The mandatory use of zone closers arises from the fact that the GELLMU system is not monolithic. Rather than being a single program, it is a suite of cross-platform component programs, each with a well-defined task, managed with a driver script. The first stage of processing operates at the level of syntax with almost no knowledge of markup vocabulary.[1] Because of this, GELLMU source, like Texinfo source, has stricter syntactic requirements than plain TeX and classical LaTeX.

Other ways in which GELLMU source differs from classical LaTeX source include:

1. Command arguments must be explicitly braced.
2. There may be no white space between a command name and the delimiter (a brace or bracket) for its first argument or option.
3. There may be no white space separating the delimiters of the successive arguments and options of a command.
4. Braces for the argument of a superscript or subscript may be omitted only if the argument is a single character.
5. The semi-colon at the end of a command name (such as `\latex;` above) indicates that the command does not introduce content. Often this type of semi-colon may be omitted, and, beyond that for most purposes `\foo;` may be regarded as shorthand for `\foo{}`.
6. The command vocabulary is somewhat different.

3 Another example

Figure 1 provides a GELLMU rendition of an example posted to the UseNet newsgroup `sci.math. research` on 29 October 2002[2] by David Madore of ENS, comparing TeX markup to MathML markup in order to illustrate the undisputed point that no

---

[1] It can be given lists of names of commands with shared syntactic properties.

[2] Message id: `apmpvn$bpb$1.repost@nef.ens.fr`

In a letter to Godfrey Harold Hardy, Sṛīṇivāsa Rāmāṉujaṉ Aiyaṅkār asserts that

$$\frac{1}{1+\frac{e^{-2\pi\sqrt{5}}}{1+\frac{e^{-4\pi\sqrt{5}}}{1+\frac{e^{-6\pi\sqrt{5}}}{\cdots}}}} = \left(\frac{\sqrt{5}}{1+\sqrt[5]{5^{3/4}\left(\frac{\sqrt{5}-1}{2}\right)^{5/2}-1}} - \frac{\sqrt{5}+1}{2}\right)e^{2\pi/\sqrt{5}}$$

Figure 1

Figure 2

author would ever regularly want to write MathML directly.[3]

This is the GELLMU source underlying figure 1:

```
\macro{\=}{\ovbar}
\macro{\.}{\ovdot}
\newcommand{\b}[1]{\unbar{#1}}
In a letter to Godfrey Harold
Hardy, S\b{r}\={\i}\b{n}iv\={a}sa
R\={a}m\={a}\b{n}uja\b{n}
Aiya\.{n}k\={a}r asserts that
\[ \frac{1
 }{1+\frac{e^{-2\pi\sqrt{5}}
 }{1+\frac{e^{-4\pi\sqrt{5}}
 }{1+\frac{e^{-6\pi\sqrt{5}}
 }{\ldots}}}}
 =
\bal{\frac{\sqrt{5}
 }{
 1+\sqrt[5]{5^{3/4}
 \bal{
  \frac{\sqrt{5}-1}{2}
 }^{5/2}-1}}
-\frac{\sqrt{5}+1}{2}}
 e^{2\pi/\sqrt{5}} \]
```

In this markup, note first the use of GELLMU's `\macro` facility to provide emulation of classical LaTeX algorithmic accents with names that are not formed with letters. Further note the use of `\bal{ ... }` in place of the LaTeX usage `\left( ...`

`\right)`. GELLMU has various balancers of this type and will eventually have more. This is related to the fact that the markup is simply a "front" for an SGML document type where a name is needed. The processor for XHTML + MathML output will not tolerate unbalanced balancing characters in a math zone except as provided through these balancers and also through the list generator

`\vect[...]{...}{...} ... {...}` .

The kinds of weak enforcement of mathematical semantics represented by such balancing provisions and by the requirement for explicit ending of *sum*, *int*, and *prod* containers is a prelude to future optional incorporation of stronger mathematical semantics in the markup.

Madore is correct in suggesting that one doesn't want to look at the MathML markup for this, but the rendering by *Firefox*, somewhat enlarged, is captured in the screenshot that is figure 2.

4 The importance of XHTML + MathML

There are several reasons why it is important to have articles and course materials with mathematical content online in modern HTML, i.e., XHTML + MathML.

4.1 Public relations for mathematics

- To a young person XHTML + MathML represents "math on the web".
- It's more flexible and more convenient for online reading than PDF — doubly so by comparison with double-column online PDF.

---

[3] Madore ends his posting as follows: "And, to remain fully on topic, I ask: has this remarkable statement by Ramanujan ever been proven rigorously? And, if so, how complicated is it?"

William F. Hammond

The following identity may be regarded as a formulation of the Weierstrass product for the Gamma function.

$$\int_0^\infty t^x e^{-t} \frac{dt}{t} = \frac{1}{x} \prod_{k=1}^{\infty} \frac{\left(1 + \frac{1}{k}\right)^x}{\left(1 + \frac{x}{k}\right)}$$

Understanding the derivation of this identity is reasonable for a bright student of first year undergraduate calculus in the United States.

Figure 3

- In many cases print journals are disappearing as librarians strive to conserve money and shelf space.
- In an electronic library browsing HTML is the analogue of browsing in the stacks, while printing a PDF document is the analogue of making a copy of an article.

4.2 Special needs

- It's great for proof reading. (Enlarge and shorten the lines.)
- **Large print editions** at no cost.

  The small gamma bit presented earlier may easily be made to look like figure 3 in *Firefox* (a screenshot).
- Articles presented in XHTML + MathML comply with web accessibility guidelines. PDF documents normally do not. (In the GELLMU production stream (see figure 5) an intermediate stage file with suffix `.zml` generated during the final stage of translation to XHTML + MathML may be of more interest than the XHTML + MathML form to those wishing to generate specific output formats for various accessibility-related purposes.)

5 Compiling an article

5.1 Acquiring the software

GELLMU is based on cross-platform free software licensed under the GNU GPL. Its package is available from CTAN [2] and from the GELLMU web site [6]. The package requires several other free cross-platform programs: GNU *Emacs*, *perl*, and two standard libraries of SGML/XML software, Open SP (for *onsgmls*) and Expat (for *xmlwf*). Whatever image manipulation software is used for handling graphic inclusions in TEX on one's platform should suffice except one should note that neither PDF nor encapsulated PostScript (usually `.eps` files) may be included as an image within an HTML page. Therefore, if one is using incorporated images, one will want to have the ability to generate copies in, for example, the PNG and JPEG formats.

Linux: The required packages are generally part of a full GNU/Linux distribution. GELLMU should be installed in `/usr/local/gellmu` and symlinks to driver scripts should be made from a suitable place in one's command path.

Mac OS X and other Unix variants: The only difference from GNU/Linux is that some of the supporting packages may need to be acquired and installed.

MS Windows: The best strategy is to install a *full* Cygwin (`http://www.cygwin.com/`) distribution. Then proceed as with Linux. (It is possible to operate natively, but the author has not done so since 2002, and no native MS-Windows user has offered to update the native MS-Windows batch files.)

```
\documenttype{article}
\title{}
\begin{document}


The following identity may be regarded as a formulation of the
Weierstrass product for the Gamma function.
\[ \int_{0}^{\infty} t^x e^{-t} \frac{dt}{t} \int:
   = \frac{1}{x}
     \prod_{k=1}^{\infty}
        \frac{\bal{1 + \frac{1}{k}}^x}{\bal{1 + \frac{x}{k}}}
     \prod: \]
Understanding the derivation of this identity is reasonable for
a bright student of first year undergraduate calculus in the
United States.

\end{document}
```

<center>Figure 4</center>

5.2 Procedure

Let's package the preceding Weierstrass product markup segment as a tiny article. Because it is GELLMU, not LATEX, it begins with

<center>\documenttype{article}</center>

rather than with

<center>\documentclass{article} .</center>

See figure 4.

Beyond early-stage syntactic processing the system requires that there be a *title* in the preamble of every *article*. An empty *title* is allowed.

Text normally must be in paragraphs. (There are exceptions.) Therefore, the blank line after `\begin{document}` is essential.

It is sometimes said about LATEX that a blank line *ends* a paragraph. However, in GELLMU a blank line *begins* a paragraph.

We place the tiny article text in a file named `gammabit.glm`, with `.glm` the canonical suffix for a GELLMU source file, enter the command

```
mmkg gammabit
```

and prepare to read the scroll. At the end when all goes well there are the following outputs:

- XHTML + MathML — the best online version
- PDF — for widely distributable print
- DVI — for TEXies
- classical HTML — for challenged browsing

Additionally one might note that some level of rendering based on *cascading style sheets* (CSS) is possible for the author-level XML.

In order to understand the scroll one needs to understand the system design.

6 System components

Regular GELLMU is a system assembled from modular components. Each step along the way produces an intermediate stage output that has its own sense and that, when things go wrong, provides opportunity both for diagnosis and intervention. A flow chart for regular GELLMU is found in figure 5.

In figure 5 what I call the "side door" is the second row entry showing the possibility of translation from source languages other than the markup of regular GELLMU into the author-level XML document type corresponding to the source markup of regular GELLMU.

One will note from the scroll that SGML/XML validation is done at several stages. This validation can be important for catching the author's mistakes. When there are error messages, it is possible and important to consult the scroll's last message regarding the stage of processing. Note in this regard that the translation from elaborated XML to XHTML + MathML takes place in three stages that are not shown on the chart but that may be seen in the example scroll.

7 Further information

Much more information may be found in the *User Guide* [3], the *GELLMU Manual* [4], and the web site for GELLMU,

<center>http://www.albany.edu/~hammond/gellmu/ .</center>

A link to an online version of this document with live links should be available at the GELLMU web site for several years.

William F. Hammond

| GELLMU source | → | SGML |

↓

| Outside SGML or XML source | → | Author-level XML |

↓

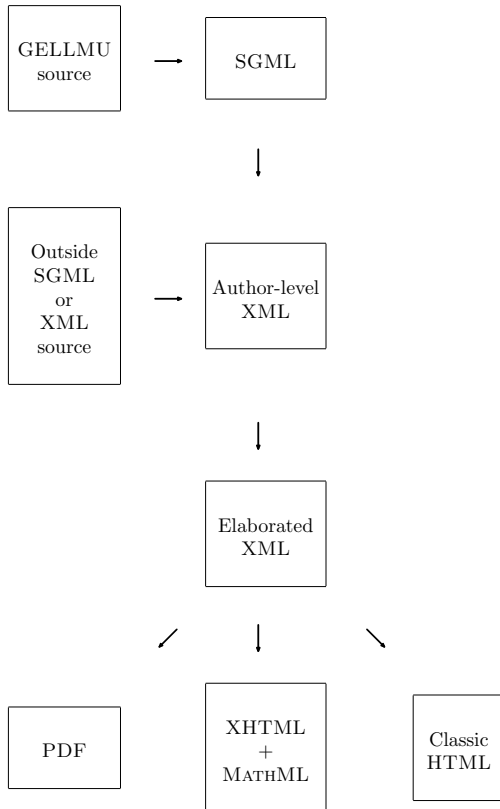| Elaborated XML |

↙ ↓ ↘

| PDF | | XHTML + MathML | | Classic HTML |

Figure 5

References

[1] William F. Hammond, "GELLMU: A Bridge for Authors from LaTeX to XML", *TUGboat: The Communications of the TeX Users Group*, vol. 22 (2001), pp. 204–207; also available online at `http://www.tug.org/TUGboat/Contents/contents22-3.html`.

[2] GELLMU at CTAN: `http://www.tex.ac.uk/tex-archive/help/Catalogue/entries/gellmu.html`

[3] William F. Hammond, "Introductory User's Guide to Regular GELLMU", `http://www.albany.edu/~hammond/gellmu/igl/userdoc.xhtml` (parallel PDF).

[4] William F. Hammond, "The GELLMU Manual", `http://www.albany.edu/~hammond/gellmu/glman/glman.xhtml` (parallel PDF).

[5] "New York Journal of Mathematics Articles in Mathematically-Capable HTML"; demonstration versions of past articles from *The New York Journal of Mathematics* ported from classical LaTeX using GELLMU, `http://math.albany.edu/demos/nyj/`.

[6] The GELLMU web site: `http://www.albany.edu/~hammond/gellmu/`