Dual Presentation with Math from One Source using GELLMU

**Preliminary Version of Talk for
the TUG Meeting in San Diego, July 2007**

William F. Hammond

Department of Mathematics & Statistics
University at Albany
Albany, New York 12222 (USA)
Email: `hammond@math.albany.edu`
Web: `http://www.albany.edu/~hammond/`

1 Introduction

A contemporary author writing an article for "dual presentation" has in mind both the classical printed presentation of an article and the modern web form of an article based on HTML.

There are two main approaches for achieving dual presentation that are relevant to the TEX community.[1]

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

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

Both methods present challenges to authors who have been accustomed to using LaTeX.

Since mid-2002 the second-generation form of HTML that supports mathematical content 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 production.

This talk will address the use of "generalized LaTeX" to produce dual content from a single LaTeX-like source. This method combines the reliability of XML document transformation with many of the conveniences available when writing LaTeX markup.

2 Writing Source Markup

Here's the source for a relatively simple example.

---

[1] *Texinfo*, the language of the GNU Documentation System, also provides a route for dual presentation of articles without mathematical markup.

```
    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.
```

It compiles to this:

> 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 LaTeX. In fact, except for the use of the zone closers `\int:` and `\prod:`, it would be LaTeX. I call it *generalized LaTeX*.

Aside from the zone closers, which are required for GELLMU, there are a few other things to notice. Most of these have to do with the fact that, as part of the overall system design, markup semantics are independent of the command vocabulary. Thus, for example:

1. Command arguments must be explicitly braced.

2. Braces for the argument of a superscript or subscript may be omitted only if the argument is a single character.

3. 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{}`.

4. The command vocabulary differs somewhat from that of LaTeX.

3 Another Example

There follows a GELLMU version of an example posted to the UseNet newsgroup `sci.math.research` on 29 October 2002, message id: `<apmpvn$bpb$1.repost@nef.ens.fr>`[2], by David Madore of ENS for the purpose of comparing TeX markup to MathML.

> In a letter to Godfrey Harold Hardy, Srīnivāsa Rāmānujan Aiyaṅkār asserts that
> $$\cfrac{1}{1+\cfrac{e^{-2\pi\sqrt5}}{1+\cfrac{e^{-4\pi\sqrt5}}{1+\cfrac{e^{-6\pi\sqrt5}}{\cdots}}}} \;=\; \left(\frac{\sqrt5}{1+\sqrt[5]{5^{3/4}\left(\frac{\sqrt5-1}{2}\right)^{5/2}-1}} - \frac{\sqrt5+1}{2} \right) e^{2\pi/\sqrt5}$$

---

[2]URI: http://groups.google.com/group/sci.math.research/msg/3bebf506912cf426

The markup for this:

```
\macro{\=}{\ovbar}
\macro{\.}{\ovdot}
\newcommand{\b}[1]{\unbar{#1}}
There follows a \abbr{GELLMU} version of an example posted to the
UseNet newsgroup \quostr{sci.math.research} on 29 October 2002, message id:
\href{http://groups.google.com/group/sci.math.research/msg/3bebf506912cf426}{
\quostr{<apmpvn\$bpb\$1.repost@nef.ens.fr>}}, by David
Madore of ENS for the purpose of comparing \tex; markup to
\abbr{MathML}.

\begin{quotation}
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
\[
\frac{1
}{1+\frac{e^{-2\pi\sqrt{5}}
}{1+\frac{e^{-4\pi\sqrt{5}}
}{1+\frac{e^{-6\pi\sqrt{5}}
}{\ldots}}}}
=
\bal{\frac{\sqrt{5}
}{
1+\sqrt[5]{5^{3/4}\bal{\frac{\sqrt{5}-1}{2}}^{5/2}-1}}
-\frac{\sqrt{5}+1}{2}}
e^{2\pi/\sqrt{5}}}
\]
\end{quotation}
```

Note the use of \bal{ ... } instead of the LATEX usage \left(... \right). GELLMU
has various balancers of this type and will eventually have more. This is related not only
to the fact that the markup is simply a "front" for an SGML document type but also to its
notion of sound mathematical semantics. 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

$$\text{\vect[...]\{...\}\{...\} ... \{...\} .}$$

The kinds of weak enforcement of mathematical semantics represented by such balancing pro-
visions and by the requirement for explicit ending of *sum*, *int*, and *prod* containers is 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 this screenshot:

3

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 fits the paradigm of web browsing.

- It's more flexible and more convenient for online reading than PDF — doubly so by comparison with double-column online PDF.

- The journals are disappearing from our libraries.

4.2 Special Needs

- It's great for proof reading. (Enlarge and shorten the lines.)

- Articles presented in XHTML+MathML comply with web accessibility guidelines. (PDF with math does not.)

- **Large print editions** at no cost.
  The small gamma bit presented earlier may easily be made to look like this in *Firefox* (a screenshot):

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.

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 a number of other cross-platform free software packages including *GNU Emacs*, *Perl*, and some standard items of SGML/XML software, chiefly *Open SP*.

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

MacOS-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 acquire and install a **full** *Cygwin*[3] distribution. Then proceed as with *Linux*.

5.2 Procedure

Let's package the preceding markup segment as an article. Because it is GELLMU, not LaTeX, it begins with \documenttype rather than with \documentclass.

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

---

[3]URI: http://www.cygwin.com/

```
    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}
```

Beyond early-stage syntatic 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 this 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[4]. At the end when all goes well there are the following outputs:

- XHTML+MATHML – the best online version[5]

- PDF – for widely distributable print[6]

- DVI – for TeXies[7]

- classical HTML – for challenged browsing[8]

Additionally one might note that some level of rendering based on *cascading style sheets (*CSS*)* is possible for the author-level XML[9].

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.

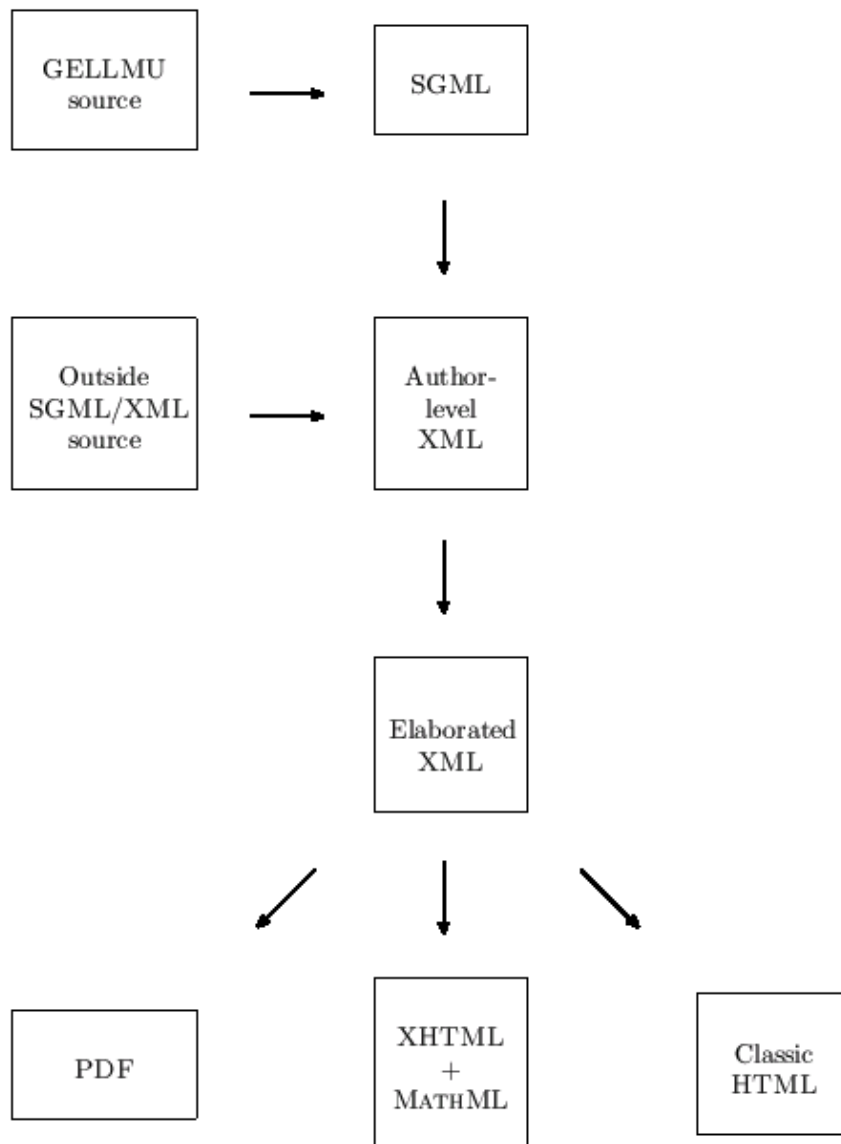[4]URI: gammabit.out
[5]URI: gammabit.xhtml
[6]URI: gammabit.pdf
[7]URI: gammabit.dvi
[8]URI: gammabit.html
[9]URI: gammabit.xml

Here's the flow chart:

```
┌──────────┐              ┌──────────┐
│ GELLMU   │    ───▶      │  SGML    │
│ source   │              │          │
└──────────┘              └──────────┘
                               │
                               ▼
┌──────────┐              ┌──────────┐
│ Outside  │              │ Author-  │
│ SGML/XML │    ───▶      │  level   │
│ source   │              │  XML     │
└──────────┘              └──────────┘
                               │
                               ▼
                          ┌──────────┐
                          │Elaborated│
                          │   XML    │
                          └──────────┘
                        ↙     │     ↘
                ┌──────────┐┌──────────┐┌──────────┐
                │          ││  XHTML   ││          │
                │   PDF    ││    +     ││ Classic  │
                │          ││ MathML   ││  HTML    │
                └──────────┘└──────────┘└──────────┘
```

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 documenttype corresponding to the source markup of regular GELLMU.

One will note from the scroll[10] 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 3 stages that are not shown on the chart but that may be seen in the example scroll.

---

[10]URI: gammabit.out

7 Further Information

Much more information may be found in the *User Guide* ([3]), the GELLMU *Manual* ([4]), and the GELLMU web site, `http://www.albany.edu/~hammond/gellmu/`. A link to an online version of this preliminary document should be available at the GELLMU web site by July 6.

References

[1]   William F. Hammond, "GELLMU: A Bridge for Authors from LaTeX to XML", *TUG-Boat: 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` (PDF[11]).

[4]   William F. Hammond, "The GELLMU Manual", `http://www.albany.edu/~hammond/gellmu/glman/glman.xhtml` (PDF[12]).

[5]   "New York Journal of Mathematics Articles in Mathematically-Capable HTML"[13]; demonstration versions of past articles from *The New York Journal of Mathematics* ported from classical LaTeX using GELLMU.

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

---

[11]URI: http://www.albany.edu/~hammond/gellmu/igl/userdoc.pdf
[12]URI: http://www.albany.edu/~hammond/gellmu/glman/glman.pdf
[13]URI: http://math.albany.edu/demos/nyj/