Generating Correct Mathematical Documents

William F. Hammond

September, 2001
Mathematical Articles

• Contain ordinary text and mathematical content.

• Need realization on paper and on computer networks.
The \textit{gamma function} may be defined in a suitable right-half plane by an integral, which essentially amounts to Fourier transform relative to the multiplicative group of postive real numbers of the reciprocal exponential function, and then extended to a meromorphic function in the complex plane with entire reciprocal. The following formula represents a variant of its Weierstrass product expansion:

\[ \int_0^\infty t^x e^{-t} \frac{dt}{t} = \frac{1}{x} \prod_{k=1}^\infty \frac{(1 + \frac{1}{k})^x}{(1 + \frac{x}{k})}. \]

The product manifests simple poles at zero and each of the negative integers.
Goals

From a Single Source:

1. A typeset version of high quality for paper preprints.

2. An online version for a web preprint.

3. A typeset version of high quality meeting the requirements of a journal chosen after the time of writing.
Goals for Online Versions

1. catalogable.

2. universally accessible.

3. re-scalable and re-sizable.

4. searchable for mathematical content.

5. “clippable” for mathematical content.
Mathematical Mainstream
1980–2000

Donald Knuth’s \TeX

\LaTeX
Plain \TeX
Other \TeX Variants:
\texttt{AMSTeX}, Con\TeXt, \textit{Omega}, \texttt{Texinfo}
How May One Provide Multiple Formattings?

Early Ideas:

1. Intuitive Authoring Systems: the WYSIWYG Idea

2. Write HTML and Translate From There.

3. Write \LaTeX and Translate From There.
WYSIWYG: The Good News

- Standard with Common “Word Processors”

- Available for \LaTeX.

- Available for SGML Systems.

- Easy for Easy Tasks.

- OK for Very Simple Documents.
WYSIWYG: The Bad News

• Hard for Hard Tasks.

• Slow for Those Who Write Frequently.

• Insufficiently Rich for Mathematics.

• Insufficiently Abstract.

• Inapplicable to Multiple Formattings.

• Some say WYSIAYG: “what you see is all you get”.
WYSIWYG vs. Structured Markup

- Format-specific hanging indentation commands.

- Use of abstract list structures.
Translating HTML

- Reliable

- But:

  1. No Math in HTML
  2. HTML Generally Less Rich Than \LaTeX
  3. Special characters are translation headaches.
    Examples: \# \$ \% \& \~ \_ \^ \\ \{ \} \< \>
Translating $\LaTeX$

- Almost Impossible
- Good Structure a Help
- May Require Human Intervention
- Need to Proof Read Twice
- $htlatex$ in the $\TeX$Live 6 distribution is remarkably good.
XML

eXtensible Markup Language

- Data Under a Template for Translation.
- Enforced Separation of Content and Presentation.
- Universal Exchange.
- Originated by

  World Wide Web Consortium (W3C)
  Sun Microsystems
XML

• Many Templates

• Synonym for XML Template:

  Document Type

• Two worlds

  1. Classical Documents:
     Examples: HTML, Docbook, TEI, ...

  2. Electronic Data Interchange (EDI)
     Example: Graham William’s T\text{E}\text{X} Catalogue found on CTAN
     help/Catalogue/catalogue.html
The GELLMU Project

• Superseded my earlier ideas:
  – Strictly controlled dialect of \LaTeX{}.
  – Adapting Texinfo (already suitable for multiple formattings).

• Relation to the Goals:
  – No present full realization of online version goals.
  – Proof that full realization of all goals and more is possible.
GELLMU

Generalized Extensible \LaTeX-Like Markup

- A markup interface for writing (SGML or) XML.

- \LaTeX-like notation more succinct than that of XML.

- Extensible using GELLMU’s \texttt{\newcommand} with arguments. (SGML has no analogue of macros with arguments.)
Difference Between \LaTeX Source and GELLMU Source

- Article prepared under a template for processing by \textit{latex}, the Program.

- Article prepared under a template for processing by many programs.

It is a small step from \LaTeX source to GELLMU source.
The Idea of LaTeX-Like MarkUp

- Text + Commands

- A *command* is a function that operates on text.

- A command may take a non-negative number of arguments.
Examples of \LaTeX-Like Commands

Example of a command taking no argument:

\latex
\LaTeX

Example of a command taking one argument:

this is \textit{emphasized} text
this is \textit{emphasized} text

Example of a command (for math) taking two arguments:

\[
\frac{az + b}{cz + d}
\]
GELLMU Modes

1. Basic

2. Advanced
   
   (a) Regular

   (b) Other (less fully developed)
Regular GELLMU: System Stages

1. \LaTeX-Like Source.

2. Syntactic Translation to SGML.

3. Translation of SGML to Enriched XML.

4. Various Formatteings of Enriched XML.
Customizing

• Each stage presents opportunities for customizing.

• Each output format is the result of successive transforms.

• Additional intermediate transforms can be provided.

• These slides were prepared using a special formatting of regular GELLMU’s article.

• The source markup* for these slides is as readable as ordinary LaTeX source.

*URI: correct.glm
The Syntactic Translator

source markup  →  XML or SGML

\foo{ ... }  →  <foo> . . . </foo>
\foo;       →  <foo/>
\foo        →  <foo>
\foo:       →  </foo>
\foo[a="x" ...]  →  <foo a="x" ...>
Syntactic Differences from \LaTeX

- Command names (element names) may contain numbers.

- Example: \texttt{\frac23} is a command name.

- Arguments must be delimited with braces or brackets.

- No white space between command name and first argument delimiter.

- No white space between delimiters of successive arguments.

- Bracketed arguments may not be optional.
Use of GELLMU in Basic Mode for XHTML

Write:

the WWW \a[href="http://www.w3.org/"]{Consortium} site

for generating the XML:

the WWW <a href="http://www.w3.org/">Consortium</a> site

to produce:

the WWW Consortium* site

*URI: http://www.w3.org/
\newcommand with XHTML

Definitions
\newcommand{\emph}[1]{\em{#1}}
\newcommand{\w3ref}[2][]{%
a[href="http://www.w3.org/#1"]{#2}}

Invocations

Using GELLMU’s $\textbf{newcommand}$ one can reduce the markup required for an anchor to \w3ref{W3C}’s \w3ref{Math/}{MathML} site.

Rendering: Using GELLMU’s \textit{newcommand} one can reduce the markup required for an anchor to W3C’s MathML site.
Why is GELLMU’s article “Didactic”? 

• Intended as a first XML document type for \LaTeX{} authors

• Sits in the middle between

  1. What \LaTeX{} authors are accustomed to.

  2. What high end XML people think is needed.

• Room to adjust and expand.
The Gamma Function: Its Weierstrass Product

\[ \int_0^\infty t^x e^{-t} \frac{dt}{t} = \frac{1}{x} \prod_{k=1}^{\infty} \frac{(1 + \frac{1}{k})^x}{(1 + \frac{x}{k})} \]
Regular GELLMU source for the identity:

\[
\int_0^\infty t^x e^{-t} \frac{dt}{t} = \frac{1}{x} \prod_{k=1}^{\infty} \frac{1 + \frac{1}{k}}{1 + \frac{x}{k}}
\]
\[ \int_0^\infty t^x e^{-t} \frac{dt}{t} = \frac{1}{x} \prod_{k=1}^{\infty} \frac{1 + \frac{1}{k}}{x} \]
\[
\frac{x}{k}
\]
Gamma: in MathML
(not by automatic translation)

<math
xmlns="http://www.w3.org/1998/Math/MathML"
class="display" mode="display">
<mrow>
<msubsup>
<mo>&Integral;</mo>
mrow>
<mrow><mn>0</mn></mrow>
<mi>&infin;</mi>
</msubsup>
<mrow>
<msup>
mrow><mi>t</mi>
mrow>
<mi>x</mi>
</msup>
<msup>
mrow><mi>e</mi>
mrow>
<mi>-t</mi>
</msup>
</mrow>
<mo> </mo>
<mfrac>
mrow><mi>dt</mi>
mrow>
<mi>t</mi>
</mfrac>
</mrow>
<mo>=</mo>
<mrow>
<mfrac>
mrow><mn>1</mn>
mrow>
<mi>x</mi>
</mfrac>
</mrow>
</mrow></math>
\[
\prod_{k=1}^{\infty} \frac{\frac{1}{1 + \frac{1}{k}}}{\frac{1}{1 + \frac{x}{k}}}
\]
Viewing MathML

Viewing support for MathML in web pages is not yet widely available. The above item can be rendered by:

- W3C’s Amaya: wprod.html or wprod.xml.

- Mozilla’s MathML development track: wprod.xml (only).

- With special plugin for MSIE: wprod.html (only).
Generating MathML from article

• Ad hoc `wprod.html` was made from GELLMU source:
  `wprod.glm`.

• The short `article` form (slide 28) of GELLMU source above `could` be given automatic translation to MathML.

• An automatic translation should go through `content` MathML and from there to `presentation` MathML.

• An automatic translation would not be under the umbrella of general XML processing.
Reliable Generation of MathML

Reliable translation will require:

A substantial non-XML, but XML-aware, parsing of all math zones in a GELLMU source document.

Occasional math parsing hints from authors in their markup.

Desirable, sometimes required:
1. Source markup labeling of math symbols.
2. Source markup typing of math symbols.
MathML Generation Issues

- Will authors cooperate?

- Will standard web user agents cooperate?
Two Final Notes

• For more information: http://www.albany.edu/~hammond/gellmu/

• GELLMU source for these slides is on the web:
  http://math.albany.edu:8000/~hammond/Presen/Correct/correct.glm