Generating Correct Mathematical Documents

William F. Hammond

September, 2001

1 Mathematical Articles

- Contain ordinary text and mathematical content.
- Need realization on paper and on computer networks.

Comment

What are our requirements for mathematical articles?

An example will indicate the scope of the task.

2 A Segment of an Article

The *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:

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

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

Comment

What are our goals for article preparation?

3 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.

4 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.

Comment

How may an article containing a segment like that of the previous example be prepared?

5 Mathematical Mainstream 1980–2000

Donald Knuth's TEX

LATEX

Plain TEX

Other TEX Variants:

AMSTEX, ConTEXt, *Omega*, *Texinfo*

Comment

LATEX is the most widely used TEX variant.

Texinfo is the TEX variant most suited to multiple presentations from a single source, but it is limited in regard to mathematics

Context by Hans Hagen is new and worth a look. It was the basis of an astounding graphics presentation at TUG 2001.

Context and *latex* both provide ways to write formatters for XML documents.

6 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.

Comment

WYSIWYG stands for "what you see is what you get". But what one sees on a screen may not be the same as what one sees on paper. This issue is compounded for source documents that admit multiple formattings.

A standard answer to those who inquire about WYSIWYG is that WYSIWYG really means WYSIAYG: "what you see is all you get".

7 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.

8 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".

9 WYSIWYG vs. Structured Markup

- Format-specific hanging indentation commands.
- Use of abstract list structures.

10 Translating HTML

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

11 Translating LaTeX

- Almost Impossible
- Good Structure a Help
- May Require Human Intervention
- Need to Proof Read Twice
- *htlatex* in the TeXLive 6 distribution is remarkably good.

12 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

13 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 TEX Catalogue found on CTAN
     `help/Catalogue/catalogue.html`

14 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.

15 GELLMU

Generalized **E**xtensible LATEX-**L**ike **M**ark**U**p

- A markup interface for writing (SGML or) XML.
- LATEX-like notation more succinct than that of XML.
- Extensible using GELLMU's *newcommand* with arguments. (SGML has no analogue of macros with arguments.)

16 Difference Between LATEX Source and GELLMU Source

- Article prepared under a template for processing by *latex*, the Program.
- Article prepared under a template for processing by many programs.

It is a small step from LATEX source to GELLMU source.

17 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.

18 Examples of LaTeX-Like Commands

Example of a command taking no argument:

```
\latex
```
LaTeX

Example of a command taking one argument:

```
this is \emph{emphasized} text
```
this is *emphasized* text

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

```
\[ \frac{a z + b}{cz + d} \]
```

$$\frac{az + b}{cz + d}$$

19 GELLMU Modes

1. Basic
2. Advanced
   (a) Regular
   (b) Other (less fully developed)

20 Regular GELLMU: System Stages

1. LaTeX-Like Source.
2. Syntactic Translation to SGML.
3. Translation of SGML to Enriched XML.
4. Various Formattings of Enriched XML.

21 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[1] for these slides is as readable as ordinary LaTeX source.

22 The Syntatic Translator

source markup	⟶	XML or SGML
`\foo{ ... }`	→	`<foo> . . . </foo>`
`\foo;`	→	`<foo/>`
`\foo`	→	`<foo>`
`\foo:`	→	`</foo>`
`\foo[a="x" ...]`	→	`<foo a="x" ...>`

23 Syntactic Differences from LaTeX

- Command names (element names) may contain numbers.
- Example: `\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.

24 Use of GELLMU in Basic Mode for XHTML

Write:

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

for generating the XML:

[1] URI: correct.glm

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

to produce:

the WWW Consortium[2] site

25 \newcommand with XHTML

Definitions

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

Invocations

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

Rendering: Using GELLMU's *newcommand* one can reduce the markup required for an anchor to W3C's MathML site.

26 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.

27 The Gamma Function: Its Weierstrass Product

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

[2]URI: http://www.w3.org/

28 Markup for the Gamma Identity

Regular GELLMU source for the identity:

```
\[ \int_{0}^{\infty}
      t^x e^{-t} \frac{d t}{t}
   \int:
   = \frac{1}{x}
   \prod_{k=1}^{\infty}
     \frac{
      \bal{1 + \frac{1}{k}}^x
     }{
      \bal{1 + \frac{x}{k}}
   } \prod: \]
```

29 Gamma: Derived XML Markup

```
<displaymath>
<int>
 <msub>0</msub>
 <msup><infty/></msup>
 t<pow>x</pow> e<pow><minus/>t</pow>
 <frac>
  <numr>d t</numr>
  <denm>t</denm>
 </frac>
</int>
<equals/>
<frac><numr>1</numr><denm>x</denm></frac>
<prod>
 <msub>k<equals/>1</msub>
 <msup><infty/></msup>
 <frac>
  <numr>
   <bal>1<plus/>
    <frac>
     <numr>1</numr>
     <denm>k</denm>
    </frac>
   </bal><pow>x</pow>
  </numr>
  <denm>
   <bal>1 <plus/>
    <frac>
     <numr>x</numr>
     <denm>k</denm>
    </frac>
   </bal>
```

```
  </denm>
 </frac>
</prod>
</displaymath>
```

30 Gamma: in MathML

(not by automatic translation)

```
<math
 xmlns="http://www.w3.org/1998/Math/MathML"
 class="display" mode="display">
<mrow>
 <mrow>
  <msubsup>
   <mo>&Integral;</mo>
   <mrow><mn>0</mn></mrow>
   <mi>&infin;</mi>
  </msubsup>
  <mrow>
   <msup>
    <mrow><mi>t</mi></mrow>
    <mrow><mi>x</mi></mrow>
   </msup>
   <mo> </mo>
  <msup>
   <mrow><mi>e</mi></mrow>
   <mrow><mi>-t</mi></mrow>
  </msup>
  <mo> </mo>
   <mfrac>
    <mrow><mi>dt</mi></mrow>
    <mi>t</mi>
   </mfrac>
  </mrow>
 </mrow>
 <mo>=</mo>
 <mrow>
  <mfrac>
   <mrow><mn>1</mn></mrow>
   <mi>x</mi>
  </mfrac>
 <mo> </mo>
  <msubsup>
   <mo>&Product;</mo>
   <mrow><mi>k</mi><mo>=</mo><mn>1</mn></mrow>
   <mi>&infin;</mi>
  </msubsup>
  <mrow>
   <mrow>
    <mfrac>
     <mrow>
      <msup>
       <mrow><mfenced>
        <mrow>
         <mn>1</mn><mo>+</mo>
         <mfrac><mn>1</mn><mi>k</mi></mfrac>
        </mrow>
       </mfenced></mrow>
       <mrow><mi>x</mi></mrow>
      </msup>
     </mrow>
     <mrow><mfenced>
      <mrow>
       <mn>1</mn><mo>+</mo>
       <mfrac><mi>x</mi><mi>k</mi></mfrac>
      </mrow>
     </mfenced></mrow>
    </mfrac>
   </mrow>
  </mrow>
 </mrow>
</mrow>
</math>
```

31 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).

32 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.

33 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.

34 MathML Generation Issues

- Will authors cooperate?
- Will **standard** web user agents cooperate?

35 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`

Document network location for HTML:

`http://math.albany.edu:8000/math/pers/hammond/Presen/Correct/correct.html`