GELLMU

A Bridge from LaTeX to XML

William F. Hammond

University at Albany

Albany, New York (USA)

http://www.albany.edu/~hammond

CTAN: support/gellmu

TUG 2001, August, 2001

Single Source Authoring

Single Source Authoring

Situation Wanted

Single Source Authoring

Situation Wanted

Write a single source file to produce:

Single Source Authoring

Situation Wanted

Write a single source file to produce:

- Finely Typeset Print

Single Source Authoring

Situation Wanted

Write a single source file to produce:

- Finely Typeset Print
- Consistent, Valid HTML

Single Source Authoring

Situation Wanted

Write a single source file to produce:

- Finely Typeset Print
- Consistent, Valid HTML

and to benefit from:

Single Source Authoring

Situation Wanted

Write a single source file to produce:

- Finely Typeset Print
- Consistent, Valid HTML

and to benefit from:

- No Need for Intervention Beyond Source

Single Source Authoring

Situation Wanted

Write a single source file to produce:

- Finely Typeset Print
- Consistent, Valid HTML

and to benefit from:

- No Need for Intervention Beyond Source
- The Possibility of Other Translations

Translating LaTeX

Translating LaTeX

- Almost Impossible

Translating LaTeX

- Almost Impossible
- Good Structure a Help

Translating LaTeX

- Almost Impossible
- Good Structure a Help
- May Require Human Intervention

Translating LaTeX

- Almost Impossible
- Good Structure a Help
- May Require Human Intervention
- Need to Proof Read Twice

Translating HTML

Translating HTML

- Reliable

Translating HTML

- Reliable
- But:

Translating HTML

- Reliable

- But:

  1. No Math in HTML

Translating HTML

- Reliable
- But:
  1. No Math in HTML
  2. HTML Generally Less Rich Than LaTeX

Translating HTML

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

Pie in the Sky

Pie in the Sky

A Dialect of classical LaTeX that is

Pie in the Sky

A Dialect of classical LaTeX that is

- Translatable

Pie in the Sky

A Dialect of classical LaTeX that is

- Translatable
- Rich

Pie in the Sky

A Dialect of classical LaTeX that is

- Translatable
- Rich
- Agreed Upon

XML

XML

eXtensible Markup Language

XML

eXtensible Markup Language

- Data Under a Template for Translation

XML

eXtensible Markup Language

- Data Under a Template for Translation
- Originated by

  World Wide Web Consortium (W3C)
  Sun Microsystems

XML

eXtensible Markup Language

- Data Under a Template for Translation
- Originated by

  World Wide Web Consortium (W3C)
  Sun Microsystems

- Universal Exchange

XML

XML

- Many Templates

XML

- Many Templates
- Synonym for XML Template:

  Document Type

XML

- Many Templates
- Synonym for XML Template:

  Document Type

- Two worlds

XML

- Many Templates
- Synonym for XML Template:

  Document Type

- Two worlds

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

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

GELLMU

GELLMU

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

GELLMU

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

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

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.

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.)

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.)
- Other *metacommand* facilities including:
  (1) *documenttype* (2) *macro*

Modes

Modes

1. Basic

Modes

1. Basic

2. Advanced

Modes

1. Basic

2. Advanced

   (a) Regular

Modes

1. Basic

2. Advanced

   (a) Regular

   (b) Other (less fully developed)

Basic GELLMU for XHTML

Basic GELLMU for XHTML

Source	Derived XML	Presentation
`\em{bird}`	`<em>bird</em>`	*bird*
`\b{cat}`	`<b>cat</b>`	**cat**
`\kbd{dog}`	`<kbd>dog</kbd>`	`dog`
`\hr;`	`<hr />`	(horizontal rule)

The Syntatic Translator

The Syntatic Translator

source markup $\longrightarrow$ XML or SGML

The Syntatic Translator

source markup $\longrightarrow$ XML or SGML

`\foo{ ... }`	$\rightarrow$	`<foo> . . . </foo>`
`\foo;`	$\rightarrow$	`<foo/>`
`\foo`	$\rightarrow$	`<foo>`
`\foo:`	$\rightarrow$	`</foo>`
`\foo[a="x" ...]`	$\rightarrow$	`<foo a="x" ...>`

Syntactic Differences from LaTeX

Syntactic Differences from LaTeX

- Command names (element names) may contain numbers.

Syntactic Differences from LaTeX

- Command names (element names) may contain numbers.
- Example: `\frac23` is a command name.

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.

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.

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.

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.

Syntax in Basic Mode

Syntax in Basic Mode

Miscellaneous Rules

Syntax in Basic Mode

Miscellaneous Rules

Brackets are only for attribute specifications.

Syntax in Basic Mode

Miscellaneous Rules

Brackets are only for attribute specifications.

Unescaped loose braces are insignificant.

Syntax in Basic Mode

Miscellaneous Rules

Brackets are only for attribute specifications.

Unescaped loose braces are insignificant.

Escaping in Basic Mode

Syntax in Basic Mode

Miscellaneous Rules

Brackets are only for attribute specifications.

Unescaped loose braces are insignificant.

Escaping in Basic Mode

Special character	%	\	{	}	#
Escaped form	\%	\\	\{	\}	\#

Basic GELLMU for XHTML

Basic GELLMU for XHTML

Anchors

Write:

Basic GELLMU for XHTML

Anchors

Write:

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

for generating the XML:

Basic GELLMU for XHTML

Anchors

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:

Basic GELLMU for XHTML

Anchors

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

\newcommand with XHTML

Definitions

\newcommand with XHTML

Definitions

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

\newcommand with XHTML

Definitions

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

Invocations

\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.
```

\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.

A CTAN Catalogue Entry

A CTAN Catalogue Entry

```
\begin{entry}[
  id="gellmu"
  datestamp="2001/07/30"
  modifier="hammond@math.albany.edu"
]
\begin{about}
\name{gellmu}
\caption{LaTeX-like markup for
         writing XML documents}
\author{\name{William F. Hammond}
  \email{hammond@math.albany.edu}}
\license[type="gpl"];
\version{\number{0.7.4}
  \released{2001/07/26}}
\end{about}
\begin{description}
\begin{abstract}
 . . .
\end{abstract}
\end{description}
\distribution{
 \ctan{support/gellmu}
}
\end{entry}
```

CTAN Catalogue XML

CTAN Catalogue XML

```
<entry
  id="gellmu"
  datestamp="2001/07/30"
  modifier="hammond@math.albany.edu"
>
<about>
<name>gellmu</name>
<caption>LaTeX-like markup for
         writing XML documents</caption>
<author><name>William F. Hammond</name>
  <email>hammond@math.albany.edu</email></author>
<license type="gpl"/>
<version><number>0.7.4</number>
  <released>2001/07/26</released></version>
</about>
<description>
<abstract>
 . . .
</abstract>
</description>
<distribution>
 <ctan>support/gellmu</ctan>
</distribution>
</entry>
```

Advanced GELLMU

Advanced GELLMU

- Multiple Argument/Option Syntax
  Example instance: `\frac{2}{3}` for $\frac{2}{3}$ **if** the name *frac* is provided as an element with two required sub-elements in the document type.

Advanced GELLMU

- Multiple Argument/Option Syntax
  Example instance: `\frac{2}{3}` for $\frac{2}{3}$ **if** the name *frac* is provided as an element with two required sub-elements in the document type.

- Various Short Reference Features
  Example: The use of blank lines, as appropriate in context, for new paragraphs **if** provided.

Advanced GELLMU

- Multiple Argument/Option Syntax
  Example instance: `\frac{2}{3}` for $\frac{2}{3}$ **if** the name *frac* is provided as an element with two required sub-elements in the document type.

- Various Short Reference Features
  Example: The use of blank lines, as appropriate in context, for new paragraphs **if** provided.

- Concept of *advanced* GELLMU is not fully developed.

Advanced GELLMU

- Multiple Argument/Option Syntax
  Example instance: `\frac{2}{3}` for $\frac{2}{3}$ **if** the name *frac* is provided as an element with two required sub-elements in the document type.

- Various Short Reference Features
  Example: The use of blank lines, as appropriate in context, for new paragraphs **if** provided.

- Concept of *advanced* GELLMU is not fully developed.

- Main Instance: **Regular GELLMU**, represented by GELLMU's own didactic *article* document type.

Why is article “Didactic”?

Why is article "Didactic"?

- Intended as a first XML document type for LaTeX authors

Why is article "Didactic"?

- Intended as a first XML document type for LaTeX authors
- Sits in the middle between

Why is article "Didactic"?

- Intended as a first XML document type for LaTeX authors

- Sits in the middle between

  1. What LaTeX authors are accustomed to.

Why is 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.

Why is 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.

Advanced GELLMU for article

Advanced GELLMU for article

Source	Derived XML	Presentation
`\emph{bird}`	`<emph>bird</emph>`	*bird*
`\latex;`	`<latex/>`	LaTeX
`\frac{2}{3}`	`<frac><num>2</num>` `<den>3</den></frac>`	$\frac{2}{3}$
`\label[:` `series="n"]{}`	`<label` `series="n"></label>`	(invisible)

Gamma Function: Its Weierstrass Product

Gamma Function: Its Weierstrass Product

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

Markup for the Gamma Identity

Markup for the Gamma Identity

Regular GELLMU source for the identity:

Markup for the Gamma Identity

Regular GELLMU source for the identity:

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

Gamma: Derived XML Markup

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>
```

Gamma: in MathML

Gamma: in MathML

(not by automatic translation)

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>
```

Viewing MathML

Viewing MathML

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

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`.

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).

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

Generating MathML from article

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

Generating MathML from article

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

- The short *article* form (slide 21) of GELLMU source above *could* be given automatic translation to MathML.

Generating MathML from article

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

- The short *article* form (slide 21) 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.

Generating MathML from article

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

- The short *article* form (slide 21) 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 Generation of MathML

Reliable translation will require:

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.

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.

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:

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.

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

MathML Generation Issues

- Will authors cooperate?

MathML Generation Issues

- Will authors cooperate?

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

How Were These Slides Made?

There were two sets of slides and, correspondingly, two formatters, one for transparencies formatted by the regular program *latex* and the other for PP4/PDF web slides formatted by the program *pdflatex* using a number of packages including *ppower4* by Klaus Guntermann of Darmstadt University of Technology. Actually I used a small modification of `"pp4slide.sty"` named `gpp4slide.sty` to make things work with the standard *slides* document class. Both of these work with GELLMU *article*, and there are slightly different definitions of *slide* with *newcommand* in the two.