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

1 Single Source Authoring

Situation Wanted

Write a single source file to produce:

• Finely Typeset Print
• Consistent, Valid HTML

and to benfit from:

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

2 Translating \LaTeX{}

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

3 Translating HTML

• Reliable

• But:
  1. No Math in HTML
  2. HTML Generally Less Rich Than \LaTeX{}
  3. Nuisances: # $ \% \& \~ \^ \{ \} < >
4 Pie in the Sky

A Dialect of classical \LaTeX{} that is

- Translatable
- Rich
- Agreed Upon

5 XML

eXtensible Markup Language

- Data Under a Template for Translation
- Originated by
  - World Wide Web Consortium (W3C)
  - Sun Microsystems
- Universal Exchange

6 XML

- Many Templates
- Synonym for XML Template: 
  \textit{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 
     \url{help/Catalogue/catalogue.html}

7 GELLMU

\textbf{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.)
- Other \textit{metacommmand} facilities including:
  (1) \texttt{\documenttype} (2) \texttt{\macro}
8 Modes

1. Basic

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

9 Basic GELLMU for XHTML

<table>
<thead>
<tr>
<th>Source</th>
<th>Derived XML</th>
<th>Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>\em{bird}</td>
<td>&lt;em&gt;bird&lt;/em&gt;</td>
<td>bird</td>
</tr>
<tr>
<td>\b{cat}</td>
<td>&lt;b&gt;cat&lt;/b&gt;</td>
<td>cat</td>
</tr>
<tr>
<td>\kbd{dog}</td>
<td>&lt;kbd&gt;dog&lt;/kbd&gt;</td>
<td>dog</td>
</tr>
<tr>
<td>\hr;</td>
<td>&lt;hr /&gt;</td>
<td>(horizontal rule)</td>
</tr>
</tbody>
</table>

10 The Syntactic Translator

source markup $\rightarrow$ XML or SGML

$\backslash foo\{ ... \} \rightarrow &lt;foo&gt; . . . &lt;/foo&gt$
$\backslash foo; \rightarrow &lt;foo/&gt$
$\backslash foo \rightarrow &lt;foo&gt$
$\backslash foo: \rightarrow &lt;/foo&gt$
$\backslash foo[a="x" ...] \rightarrow &lt;foo a="x" ...&gt$

11 Syntactic Differences from \LaTeX

- Command names (element names) may contain numbers.
- Example: $\backslash frac23$ 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.

12 Syntax in Basic Mode

Miscellaneous Rules

Brackets are only for attribute specifications.
Unescaped loose braces are insignificant.
Escaping in Basic Mode

<table>
<thead>
<tr>
<th>Special character</th>
<th>%</th>
<th>\</th>
<th>{</th>
<th>)</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Escaped form</td>
<td>%\</td>
<td>\</td>
<td>{</td>
<td>}</td>
<td>#</td>
</tr>
</tbody>
</table>

13 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 Consortium¹ site

14 \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 \textit{newcommand} one can reduce the markup required for an anchor to W3C’s MathML site.

15 A CTAN Catalogue Entry

```
\begin{entry}
  \id{gellmu}
  \datestamp{2001/07/30}
  \modifier{hammond@math.albany.edu}
\end{entry}
```

¹URI: http://www.w3.org/
16 CTAN Catalogue XML

```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>...
  </description>
  <distribution>
    <ctan>support/gellmu</ctan>
  </distribution>
</entry>
```

17 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.

18 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.
19 Advanced GELLMU for article

Source
\emph{bird}
\latex;
\frac{2}{3}
\label[:series="n"]{}

Derived XML
<emph>bird</emph>
<latex/>
<frac><num>2</num><den>3</den></frac>
<label series="n">invisible</label>

Presentation
bird
\LaTeX
\frac{2}{3}
(invisible)

20 Gamma Function: Its Weierstrass Product

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

21 Markup for the Gamma Identity

Regular GELLMU source for the identity:

\begin{verbatim}
[ \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})}
\end{verbatim}

22 Gamma: Derived XML Markup

<displaymath>
<int>
<msub>0</msub><msup>\infty</msup>
t^x e^{-t} \frac{dt}{t} \prod_{k=1}^\infty \frac{(1 + \frac{1}{k})^x}{(1 + \frac{x}{k})}
</int>
<equals/>
<frac><num>1</num><den>x</den></frac>
<prod><msub>k</msub><equals/><msub>1</msub></prod>
</displaymath>
23 Gamma: in MathML

(not by automatic translation)
24 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).

25 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.

26 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.

27 MathML Generation Issues

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

28 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.

Document network location: