TEX/LATEX, MathML, TeX4ht: essential tools for creating accessible documents

Jacek Polewczak

jacek.polewczak@csun.edu

Mathematics, CSU

Northridge, CA 91330

USA

The goals

- Assembly of the tools for creating (ADA/508 compliant) accessible documents, including text, math, graphics, and field oriented (chemistry, physics, law, music);

The goals

- Assembly of the tools for creating (ADA/508 compliant) accessible documents, including text, math, graphics, and field oriented (chemistry, physics, law, music);
- Delivery of platforms' specific (Linux, Mac, Windows) tools to the faculty/students through:

The goals

- Assembly of the tools for creating (ADA/508 compliant) accessible documents, including text, math, graphics, and field oriented (chemistry, physics, law, music);
- Delivery of platforms' specific (Linux, Mac, Windows) tools to the faculty/students through:
  - ISO images for CD/DVD burning available on a CSUN server,

The goals

- Assembly of the tools for creating (ADA/508 compliant) accessible documents, including text, math, graphics, and field oriented (chemistry, physics, law, music);
- Delivery of platforms' specific (Linux, Mac, Windows) tools to the faculty/students through:
  - ISO images for CD/DVD burning available on a CSUN server,
  - master copies available for CD/DVD duplication,

The goals

- Assembly of the tools for creating (ADA/508 compliant) accessible documents, including text, math, graphics, and field oriented (chemistry, physics, law, music);
- Delivery of platforms' specific (Linux, Mac, Windows) tools to the faculty/students through:
  - ISO images for CD/DVD burning available on a CSUN server,
  - master copies available for CD/DVD duplication,
  - online downloading of particular packages;

The goals

- Assembly of the tools for creating (ADA/508 compliant) accessible documents, including text, math, graphics, and field oriented (chemistry, physics, law, music);
- Delivery of platforms' specific (Linux, Mac, Windows) tools to the faculty/students through:
  - ISO images for CD/DVD burning available on a CSUN server,
  - master copies available for CD/DVD duplication,
  - online downloading of particular packages;
- Tutorials: online and retrievable (for printing) from a CSUN server.

The goals

- Assembly of the tools for creating (ADA/508 compliant) accessible documents, including text, math, graphics, and field oriented (chemistry, physics, law, music);
- Delivery of platforms' specific (Linux, Mac, Windows) tools to the faculty/students through:
  - ISO images for CD/DVD burning available on a CSUN server,
  - master copies available for CD/DVD duplication,
  - online downloading of particular packages;
- Tutorials: online and retrievable (for printing) from a CSUN server.

Note: *The presentation below provides a brief overview of the tools/ideas involved in the process of creating accessible documents.*

It is not meant as a tutorial !

The theme . . . , LaTeX

- What is LaTeX ?

The theme . . . , LaTeX

- What is LaTeX ?
  - LaTeX is a markup language (as groff/troff, html, are) for representing structured documents.

The theme . . . , LaTeX

- What is LaTeX ?
  - LaTeX is a markup language (as groff/troff, html, are) for representing structured documents.
- Very simple examples of *LaTeX* and *html* source codes:

The theme . . . , LaTeX

- What is LaTeX ?
  - LaTeX is a markup language (as groff/troff, html, are) for representing structured documents.
- Very simple examples of *LaTeX* and *html* source codes:
  - Side by side source codes for LaTeX and *html*

What is document markup ?

- In a word processor, one sees the document more or less as it will look when printed;

What is document markup ?

- In a word processor, one sees the document more or less as it will look when printed;
- Working with a *markup language* is different. For example, to emphasize the phrase `detailed description` in LaTeX source file, one has to type:
  `\emph{detailed description};`

What is document markup ?

- In a word processor, one sees the document more or less as it will look when printed;
- Working with a *markup language* is different. For example, to emphasize the phrase `detailed description` in LaTeX source file, one has to type:
  `\emph{detailed description}`;
- Markup languages have always dominated typographic work of high quality;

What is document markup ?

- In a word processor, one sees the document more or less as it will look when printed;
- Working with a *markup language* is different. For example, to emphasize the phrase `detailed description` in LaTeX source file, one has to type:
  `\emph{detailed description}`;
- Markup languages have always dominated typographic work of high quality;
- LaTeX is a markup language designed and implemented by Leslie Lamport (1994) in *A document Preparation System* , based on Donald E. Knuth's (1984) *The TeXbook* .

The theme . . . , MathML

- What is MathML?

The theme . . . , MathML

- What is MathML?
  - MathML is an application of XML for describing mathematical notations and capturing both its structure and content. It aims at integrating mathematical formulas into World Wide Web documents.

The theme . . . , MathML

- What is MathML?
  - MathML is an application of XML for describing mathematical notations and capturing both its structure and content. It aims at integrating mathematical formulas into World Wide Web documents.
  - As LaTeX, XML is a markup language for representing structured documents. However, in contrast to LaTeX, XML is ***NOT*** page layout language. Also XML is an interchange and manipulation interface designed for machine.

The theme . . . , MathML

- What is MathML?
  - MathML is an application of XML for describing mathematical notations and capturing both its structure and content. It aims at integrating mathematical formulas into World Wide Web documents.
  - As LaTeX, XML is a markup language for representing structured documents. However, in contrast to LaTeX, XML is ***NOT*** page layout language. Also XML is an interchange and manipulation interface designed for machine.
- Very simple examples of *LaTeX* and *MathML* codes:

The theme . . . , MathML

- What is MathML?
  - MathML is an application of XML for describing mathematical notations and capturing both its structure and content. It aims at integrating mathematical formulas into World Wide Web documents.
  - As LaTeX, XML is a markup language for representing structured documents. However, in contrast to LaTeX, XML is ***NOT*** page layout language. Also XML is an interchange and manipulation interface designed for machine.
- Very simple examples of *LaTeX* and *MathML* codes:
  - Side by side source codes for LaTeX and *MathML* (***Note:*** In the interest of brevity, a preamble part of the XML/MathML document is omitted.)

Presentation and Content MathML

From Wikipedia entry for MathML :

MathML deals not only with the presentation but also the meaning of formula components (the latter part of MathML is known as "Content MathML"). Because the meaning of the equation is preserved separate from the presentation, how the content is communicated can be left up to the user. For example, web pages with MathML embedded in them can be viewed as normal web pages with many browsers but visually impaired users can also have the same MathML read to them through the use of screen readers (e.g. using the MathPlayer plugin for Internet Explorer, Opera 9.50 build 9656+ or the Fire Vox extension for Firefox).

Presentation MathML

From Wikipedia entry for MathML :

Presentation MathML focuses on the display of an equation, and has about 30 elements, and 50 attributes. The elements all begin with `m` *and include token element:* `<mi>x</mi>` - *identifiers;* `<mo>+</mo>` - *operators;* `<mn>2</mn>` - *number. Tokens are combined using layout elements which include:* `<mrow>` - *a row;* `<msup>` - *superscripts;* `<mfrac>` - *fractions. The attributes mainly control fine details of the presentation. A large number of entities are available which represent letters* `&pi` (π, my addition); *symbols* `&RightArrow`; *and some non-visible character such as* `&InvisibleTimes`; *representing multiplication.*

Below, I focus only on *Presentation MathML*.

Important qualifications

- TEX/LATEX provides extremely detailed page layout. HTML/XML/MathML formats do not! They are *functional* mark-up languages and *NOT* page layout languages. Their exact rendering is not given by the document but decided by a browser, by windows size, resolution, and font selection. The results are good for browsing by not for printing.
- The only way to produce a precise page layouts is to represent documents in a page layout languages such as PDF, Postscript, or DVI.
- Furthermore, . . . *word processor formats are not suitable for transmitting printable copy, hopeless for browsing, and unreliable for archiving because of the instability of the formats.* (see, reference [9])

Prose

In addition to mathematical expressions, MathML (as an application of XML) provides all the needed support for prose documents:

- Sectional units;
- Lists;
- Tables, etc.

For example, these LaTeX and XML source codes result in the following (browser) output, or its pdf version.

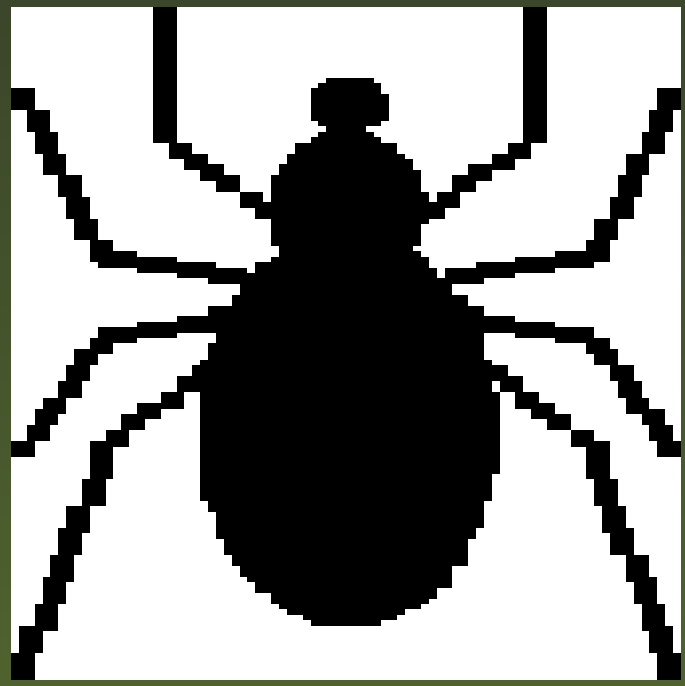
For brevity, I didn't include preambles in either LaTeX or XML codes.

Graphics

MathML (as an application) of XML easily handles pictures. Here are side by side comparisons of LaTeX and XML codes that result in the following picture:

Graphics

MathML (as an application) of XML easily handles pictures. Here are side by side comparisons of LaTeX and XML codes that result in the following picture:



Creating MathML documents

There are several ways to create MathML documents. One can start from

Creating MathML documents

There are several ways to create MathML documents. One can start from

- LaTeX document and convert the LaTeX source code to MathML using Open Source TeX4ht, TTH, Hermes, (online) ORCCA, or Tralics;

Creating MathML documents

There are several ways to create MathML documents. One can start from

- LaTeX document and convert the LaTeX source code to MathML using Open Source TeX4ht, TTH, Hermes, (online) ORCCA, or Tralics;
- one can use MathML authoring tools, for example, commercial applications like MathType and Scientific Word, or Open Source and full fledged word-processor OpenOffice.

Creating MathML documents

There are several ways to create MathML documents. One can start from

- LaTeX document and convert the LaTeX source code to MathML using Open Source TeX4ht, TTH, Hermes, (online) ORCCA, or Tralics;
- one can use MathML authoring tools, for example, commercial applications like MathType and Scientific Word, or Open Source and full fledged word-processor OpenOffice.
- One can also use MathML editors/browsers, e.g., commercial WebEQ^{TM} or Open Source Amaya.

Creating MathML documents

There are several ways to create MathML documents. One can start from

- LaTeX document and convert the LaTeX source code to MathML using Open Source TeX4ht, TTH, Hermes, (online) ORCCA, or Tralics;
- one can use MathML authoring tools, for example, commercial applications like MathType and Scientific Word, or Open Source and full fledged word-processor OpenOffice.
- One can also use MathML editors/browsers, e.g., commercial WebEQ^{TM} or Open Source Amaya.
- Additionally, the commercial algebra systems and engineering software, e.g., *Maple*, *Mathematica*, *Matlab*, or *Mathcad*, allow saving worksheets (documents) in MathML format.

Starting from LaTeX

foo.tex $\Longrightarrow$ TeX4ht $\Longrightarrow$ MathML

Starting from LaTeX

foo.tex $\Longrightarrow$ TeX4ht $\Longrightarrow$ MathML

Conversion is obtained with the following *TeX4ht* command,

mzlatex foo.tex,

and the output is XML file, foo.xml.

Starting from LaTeX

foo.tex $\Longrightarrow$ TeX4ht $\Longrightarrow$ MathML

Conversion is obtained with the following *TeX4ht* command,

mzlatex foo.tex,

and the output is XML file, foo.xml.

Notes:

(1) There is ***NO*** need to change the original LaTeX source file before conversion to MathML.

Starting from LaTeX

foo.tex $\Longrightarrow$ TeX4ht $\Longrightarrow$ MathML

Conversion is obtained with the following *TeX4ht* command,

mzlatex foo.tex,

and the output is XML file, foo.xml.

Notes:

(1) There is ***NO*** need to change the original LaTeX source file before conversion to MathML.

(2) After applying the LaTeX commands,

latex foo.tex and *pdflatex* foo.tex,

one gets usual dvi (foo.dvi) and pdf (foo.pdf) outputs.

Other converters to MathML format

There are other converters from LaTeX to MathML (all Open Source):

Other converters to MathML format

There are other converters from LaTeX to MathML (all Open Source):

- TTH: the TEX to HTML translator ; last time updated in March 2007.

Other converters to MathML format

There are other converters from LaTeX to MathML (all Open Source):

- TTH: the TEX to HTML translator ; last time updated in March 2007.
- Tralics: a LaTeX to XML translator ; currently maintained.

Other converters to MathML format

There are other converters from LaTeX to MathML (all Open Source):

- TTH: the TEX to HTML translator ; last time updated in March 2007.
- Tralics: a LaTeX to XML translator ; currently maintained.
- Hermes - a semantic XML+MathML+Unicode e-publishing/self-archiving tool for LaTeX authored scientific articles; last updated in November 2007.

Other converters to MathML format

There are other converters from LaTeX to MathML (all Open Source):

- TTH: the TEX to HTML translator ; last time updated in March 2007.
- Tralics: a LaTeX to XML translator ; currently maintained.
- Hermes - a semantic XML+MathML+Unicode e-publishing/self-archiving tool for LaTeX authored scientific articles; last updated in November 2007.
- ORCCA, TeX/LaTeX to MathML Online Translator maintained by Ontario Research Center for Computer Algebra (ORCCA). However, the site is not always operational.

Converters to MathML format, via Pseudo-TeX (all Open Source)

- Starting from an html document, ASCIIMathML JavaScript package provides quick and easily produced math formulas by using calculator-style syntax. For other variants, see the above webpage.

Converters to MathML format, via Pseudo-TeX (all Open Source)

- Starting from an html document, ASCIIMathML JavaScript package provides quick and easily produced math formulas by using calculator-style syntax. For other variants, see the above webpage.
- The following three packages:

  Generalized Extensible LaTeX-Like Markup (GELLMU) ,

  Itex2MML (Unix platform only) , and

  WebTeX: A Markup Language for WebEQ^{TM}

  modify TEX/LATEX syntax for further conversion into MathML format.

WebTeX is worth mentioning since it used as an input to the various WebEQ^{TM} tools. However, its output can take a number of forms, including images, applets, or the equivalent MathML markup.

Selected Graphical document editors that export documents to MathML I

There are several MathML authoring tools that do not require LaTeX code to create MathML output:

- Commercial MathType equation editor for Mac/Windows platforms only. *MathType editor* works with MS Office applications (MSWord, PowerPoint, and Excel documents), as well as with QuarkXPress and Adobe inDesign layouts. It also accepts TeX/LaTeX typesetting language.

Selected Graphical document editors that export documents to MathML I

There are several MathML authoring tools that do not require LaTeX code to create MathML output:

- Commercial MathType equation editor for Mac/Windows platforms only. *MathType editor* works with MS Office applications (MSWord, PowerPoint, and Excel documents), as well as with QuarkXPress and Adobe inDesign layouts. It also accepts TeX/LaTeX typesetting language.
- Commercial Scientific Word/Notebook combines MuPad$_{Pro}$ algebra system with WYSIWYG word-processing of technical documents. Its documents can be exported to LaTeX and to XML/MathML. *Scientific Word* is available for Linux/Mac/Windows platforms.

Selected Graphical document editors that export documents to MathML II

- Open Source OpenOffice is a professional word processor freely available for Linux, Mac, and Windows platforms. *OpenOffice* exports documents not only to MathML and LaTeX , but also ***reads and writes MSWord documents!***

  As in MathType and Scientific Word, OpenOffice does not require LaTeX input.

Selected MathML editors/viewers

- W3C's Editor/Browser (started in 1996), Amaya , is a tool used to create and update documents directly on the Web, including an increasing number of XML applications, such as the XHTML family, MathML, and SVG. It allows all those vocabularies to be edited simultaneously in compound documents. Open Source *Amaya editor* is a freely available for all three platforms: Linux, Mac, and Windows. It does not require LaTeX input.

Selected MathML editors/viewers

- W3C's Editor/Browser (started in 1996), Amaya , is a tool used to create and update documents directly on the Web, including an increasing number of XML applications, such as the XHTML family, MathML, and SVG. It allows all those vocabularies to be edited simultaneously in compound documents. Open Source *Amaya editor* is a freely available for all three platforms: Linux, Mac, and Windows. It does not require LaTeX input.
- Commercial WebEQ^{TM} Developers Suite is a toolkit for creating web pages that include interactive math. *WebEQ*^{TM} is based on Java and has embedded MathML. It is platform and browser independent. Available for Linux/Mac/Windows.

Other tools for creating MathML

Below, I list examples of proprietary algebra systems and engineering calculation software that export documents to XML/MathML.

- *Maple* is a mathematical computation engine with fully integrated numerics and symbolics, for modeling, simulation, and visualization, all accessible from WYSIWYG technical documents that can be exported to LaTeX and to XML/MathML.
- *Mathematica* is another algebra system, not just for computation but for modeling, simulation, visualization, development, documentation, and deployment. As in the case of *Maple*, *Mathematica* notebooks/documents can be exported to both LaTeX and XML/MathML.

. . . continued

- *Matlab* is an advanced computational environment providing a high-level programming language and functions for algorithm development, data analysis and visualization, and numeric computation. *Matlab* serves as the basis for other *Mathworks* products, for example, *Simulink* as well as dozens of application's specific toolboxes. *Matlab* provides exporting functions to LaTeX and XML formats.
  An additional small Java class library provides *Matlab* scripts for importing Content-MathML declarations (values, vectors, matrices and function handles) into the *Matlab* workspace. The same library supports exporting *Matlab* data into Content- and Presentation-MathML.

. . . continued

- *Mathcad* is engineering calculation software that integrates standard mathematical notation, text and graphs in a single worksheet. Its worksheets can be exported to XML/MathML format.

. . . continued

- *Mathcad* is engineering calculation software that integrates standard mathematical notation, text and graphs in a single worksheet. Its worksheets can be exported to XML/MathML format.

And finally, here is a more complete list of

The W3C MathML software

LaTeX again . . .

Open Source LaTeX/TeX is a professional typesetting and publishing tool (used by major publishing houses) with extremely detailed page layout; it is free to use and/or to modify. There is elegance and efficiency when the same ASCII source file can produce different outputs; dvi, postscript, and pdf for printing/viewing or XML/MathML for accessible viewing in browsers. With the use of front-ends for LaTeX, there is no need to know many technical details of this typesetting system. Below, I provide links to three examples of Open Source front-ends that are easy to install and use. LaTeX distribution is required for their proper installation.

- LyX - The Document Processor , available for all platforms;
- GNU TeX , WYSIWYW (what you see is what you want) TeXmacs editor for scientists, available for all platforms;
- Kile – an integrated LaTeXeditor for the KDE desktop environment. KDE is available for many architectures such as PC, PowerPC (Mac for example) and SPARC.

References

[1] E. M. Gurari and S. Rahtz,
From LaTeX to MathML and Back with TeX4ht and PassiveTeX

[2] E. M. Gurari, MathML via TeX4ht and other tools

[3] M. Goossens, S. Rahtz, E. M. Gurari, R. Moore, and R. S. Sutor,
The LaTeX Web Companion: Integrating TeX, HTML, and XML

[4] M. Goossens, S. Rahtz, and F. Mittelbach,
The LaTeX Graphics Companion: Illustrating Documents with TeX and Postscript

[5] G. Grätzer, Math Into LaTeX

[6] M. Goossens, F. Mittelbach, and A. Samarin, The LaTeX Companion

[7] TUG (TeX Users Group) website

[8] *Design Science - How Science Communicates*^{TM} website

[9] TTH: the TEX to HTML translator website

[10] D. ,E. Knuth, *The TeXbook*

[11] L. Lamport, *A document Preparation System*

[12] The W3C MathML software list

Source codes for LaTeX and *html*

LaTeX

```
\documentclass{article}
\begin{document}
   Hello World!
\end{document}
```

LaTeX output

Hello World!

html

```
<html>
<head>
<title>Hello HTML
</title>
</head>
<body>
Hello World!
</body>
</html>
```

browser's rendering

Hello World!

Codes for LaTeX and *MathML*

LaTeX

```
\documentclass{article}
\begin{document}
$a^4 + 5$.
\end{document}
```

LaTeX output

$a^4 + 5$.

MathML

```
<p> <math xmlns=
"http://www.w3.org/1998/Math/MathML"
        display="inline">
  <msup>  <mi>a</mi>
    <mn>4</mn>  </msup>
  <mo>+</mo>
  <mn>5</mn>
</math>.
</p>
```

browser's rendering

$a^4 + 5$.

Prose source example of LaTeX ◀

```
\tableofcontents
  \section{Nested Lists \& A Table}
    \subsection{Nested Lists}
      Two kinds of lists:
      \begin{itemize}
        \item First item
          \begin{enumerate}
            \item Sub item
            \item Sub item
          \end{enumerate}
        \item Second item
      \end{itemize}
\subsection{A Table}
    \begin{tabular}{ccc}
      one & two & three \\
      1 & 2 & 3
    \end{tabular}
  \section{A Citation}
    See \cite{abc}.
  \begin{thebibliography}{99}
    \bibitem{abc} A bib entry.
  \end{thebibliography}
```

Prose XML source example ◀ ▶

```
<h3>Contents</h3>
<div>
   1 <a href="#sec1">Nested Lists &amp;amp; A Table</a>
   <br/> 1.1 <a href="#sec1.1">Nested Lists</a>
   <br/> 1.2 <a href="#sec1.2">A Table</a>
   <br/>2 <a href="#sec2">A Citation</a>
</div>
<h3>1 <a id="sec1"/>Nested Lists &amp;amp; A Table</h3>
<h4>1.1 <a id="sec1.1"/>Nested Lists</h4>
<p>Two kinds of lists: </p>
<ul>
  <li>First item
       <ol><li>Sub item </li><li>Sub item</li></ol>
   </li>
  <li>Second item</li>
</ul>
<h4>1.2 <a id="sec1.2"/>A Table</h4>
<div>
  <table>
    <colgroup>
      <col/>
      <col/>
```

continued. . .

```
  <col/>
    </colgroup>
    <tr>
      <td>one</td>
      <td>two</td>
      <td>three</td>
    </tr>
    <tr>
      <td>1</td>
      <td>2</td>
      <td>3</td>
    </tr>
  </table>
</div>
<h3>2   <a id="sec2"/>A Citation</h3>
<p>See [<a href="#bib">1</a>]. </p>
<h3><a id="sec3"/>References</h3>
<div>
  <p> [1] <a id="bib"/>A bib entry. </p>
</div>
```

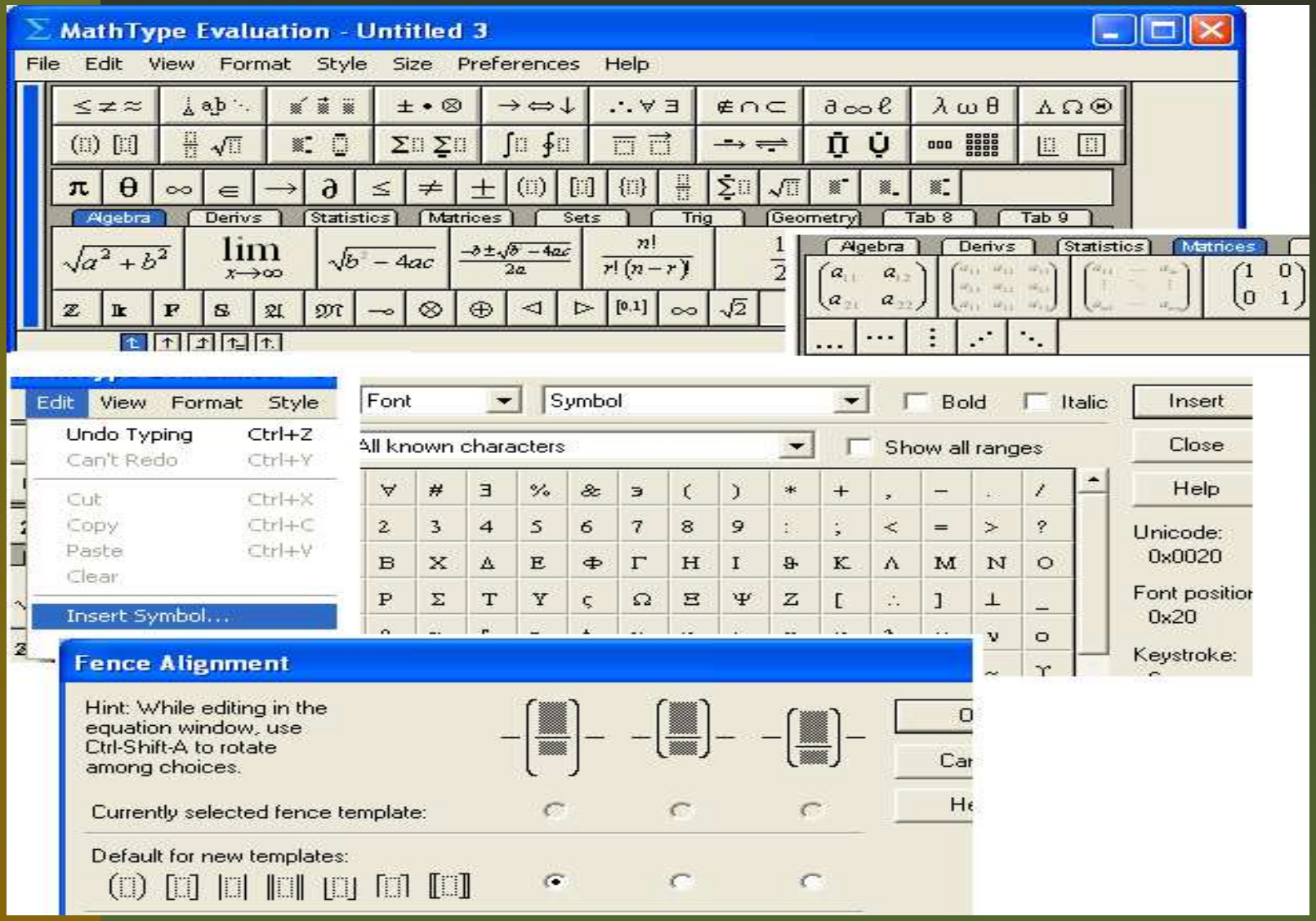
LaTeX code ◀ ▶

```
\Draw
\Scale(0.5,0.5)
{ \Line(6,26)   \Line(35,38)
                \Line(0,14) }
{ \Move(52, 0)  \Line(-6,26)
  \Line(-35,38) \Line( 0,14) }
{ \Move( 0,26)  \Line( 6,14)
  \Line( 40,9)  \Line( 6,20) }
{ \Move(52,26)  \Line(-6,14)
  \Line(-40,9)  \Line(-6,20) }
\Move(26,29) \PaintOval(12,23)
\Move(0,26)  \PaintOval(6,9)
\Move(0,12)  \PaintOval(3,3)
\EndDraw
```

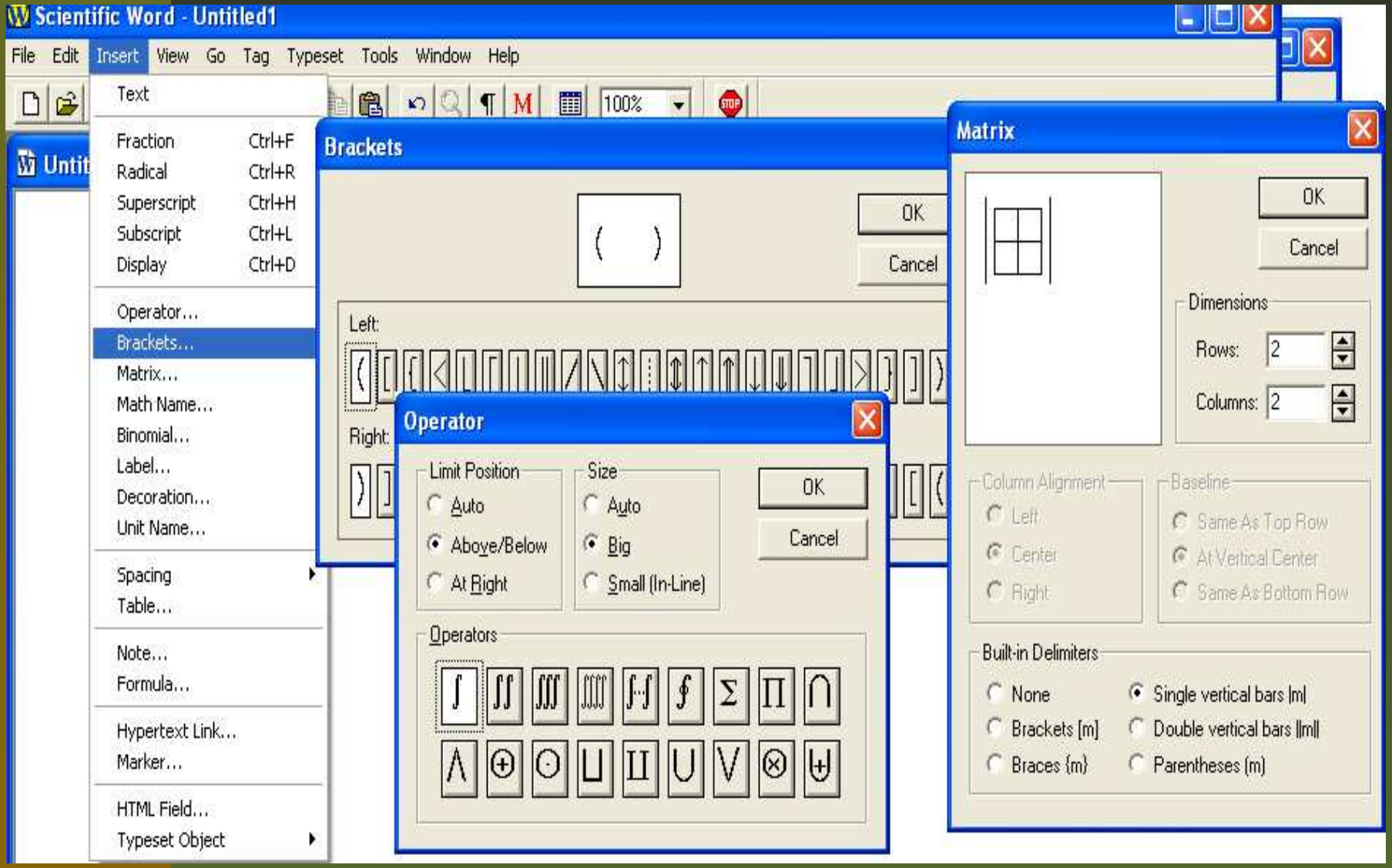
XML code ◀

```
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 1.0//EN"
"http://www.w3.org/TR/2001/REC-SVG-20010904/DTD/svg10.dtd">
<svg xmlns="http://www.w3.org/2000/svg" stroke="black">
  <polyline points="10,88 16,62 51,24 51,10" fill="none"/>
  <polyline points="62,88 56,62 21,24 21,10" fill="none"/>
  <polyline points="10,62 16,48 56,39 62,19" fill="none"/>
  <polyline points="62,62 56,48 16,39 10,19" fill="none"/>
  <ellipse cx="36" cy="59" rx="12" ry="23"/>
  <ellipse cx="36" cy="33" rx="6"  ry="9"/>
  <ellipse cx="36" cy="21" rx="3"  ry="3"/>
</svg>
```

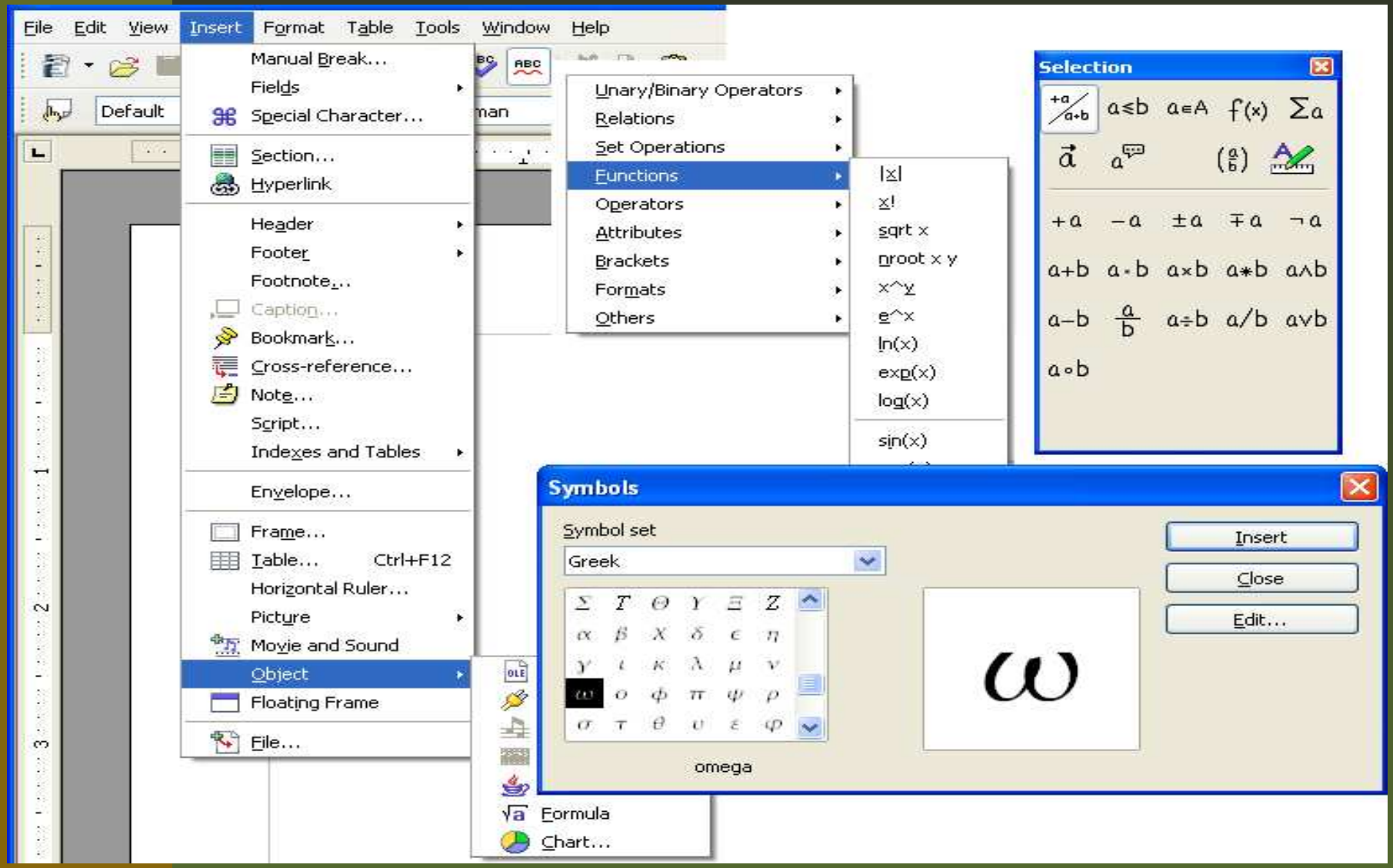
Typical screen in MathType editor



Typical screen in Scientific Word editor



Typical screen in OpenOffice editor



Typical screen in Amaya editor