MathML and other XML Technologies for Accessible PDF from LaTeX

Frank Mittelbach David Carlisle Ulrike Fischer Joseph Wright
The LaTeX Project

4th September 2025

MathML in PDF

- PDF/UA-2 mandates all mathematical content links to MathML:
  - As an Associated File (AF)
  - Using Structure Elements (SE)
- AF is easier to create/link but SE is better for synching to visual appearance
- The PDF 2.0 standard does *not* define MathML PDF Structure Elements
- But there is a working agreement for Assistive Technology (AT) systems

MathML Attributes

- *Some* MathML attributes correspond to PDF Properties
- But in general there is no simple mapping: needs agreement
- This uses a namespace approach: adds complexity

MathML from LaTeX

$$\sum_{i=1}^{n} i = \frac{(n+1)n}{2}$$

```
% !TeX program = LuaLaTeX
\DocumentMetadata{tagging = on,
  lang = en}
\documentclass{article}
\usepackage{unicode-math}
\begin{document}
\[ \sum_{i=1}^{n} i
  = \frac{(n+1)n}{2} \]
\end{document}
```

```
...
<munderover>
<mo lspace="0" movablelimits="true"
  rspace="0.167em">Σ</mo>
<mrow>
<mi>i</mi>
<mo lspace="0" rspace="0">=</mo>
<mn>1</mn>
</mrow>
<mi>n</mi>
</munderover>
...
```

Intent

What does $|x|$ mean?

Input as `\abs{x}` with suitable definition and get

```
<mrow intent="absolute-value($x)">
  <mo>|</mo> <mi arg="x">x</mi> <mo>|</mo>
</mrow>
```

Complex material and AT

For example $\sqrt{x}$ gives MathML

```
<msqrt><mi>x</mi></msqrt>
```

i.e. no rule in the MathML structure

Need to *hide* material using `Artifact` marking:
it is then ignored by AT

Validation

- PDF Structure Elements do not correspond to existing XML structures
- PDF validators do not validate MathML
- The LaTeX Team have developed a tool to extract XML representation for Structure Tree
- This can be validated using a Relax-NG Schema

Summary

- MathML embed PDF 2.0 in a usable way is becoming more established
- LaTeX can automatically include MathML in PDF output
- This enables automatic creation of accessible STEM documents
- Adjustment by the author may still be needed
- PDF Structure Trees containing custom tags can be validated:
  whether they are made by LaTeX or not