

# PreTeXt

## An XML vocabulary for scholarly documents

Balisage: The Markup Conference

---

Rob Beezer

August 3, 2018

Department of Mathematics and Computer Science

University of Puget Sound

beezer@pugetsound.edu

# Motivation

- Facts
  - Academics want to be read, so will write
  - Textbooks without commercial constraints can be better
  - Pssst! Textbooks are expensive!

# Motivation

- Facts
  - Academics want to be read, so will write
  - Textbooks without commercial constraints can be better
  - Pssst! Textbooks are expensive!
- Problem
  - Self-publishing is hard
  - Self-publishing to multiple electronic formats is harder still

# Motivation

- Facts
  - Academics want to be read, so will write
  - Textbooks without commercial constraints can be better
  - Pssst! Textbooks are expensive!
- Problem
  - Self-publishing is hard
  - Self-publishing to multiple electronic formats is harder still
- Solutions
  - Open licenses on *source*
  - Source: XML syntax, PreTeXt vocabulary
  - Processing: XSLT, assisted with Python
  - Destinations:  $\text{\LaTeX}$ , HTML+CSS, EPUB, JSON, Markdown
  - MathJax (Javascript library for mathematics in browsers)

# Output Formats

## Primary

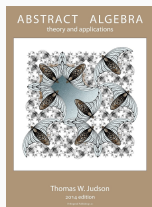
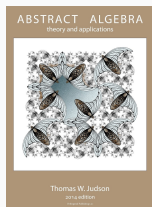
- $\text{\LaTeX}$   $\rightarrow$  PDF, for print-on-demand
- $\text{\LaTeX}$   $\rightarrow$  PDF, for medium and large screens
- HTML + CSS + MathJax + Javascript (mobile)
- EPUB (beautiful math in iBooks)
- Jupyter computational notebooks (beta)
- Nemeth Braille (prototype)
- Kindle (math is broken!)

## Subsets

- Solutions Manual
- WeBWork problem archives
- Sage unit tests
- More to come...



## ABSTRACT ALGEBRA DEMONSTRATION



# Vocabulary Features

- First priority: author-friendly

# Vocabulary Features

- First priority: author-friendly
- Very few abbreviations: **p**, **ol**, **li**, **c**, **q**, **m**
- Lower-case, few hyphens:  
**theorem**, **proof**, **subsection**, **hint**, **latex-image**
- Limited attributes, sensible defaults



# Vocabulary Features

- First priority: author-friendly
- Very few abbreviations: **p**, **ol**, **li**, **c**, **q**, **m**
- Lower-case, few hyphens:  
**theorem**, **proof**, **subsection**, **hint**, **latex-image**
- Limited attributes, sensible defaults
- Robust and unified cross-references with **xref**
- Variety of cross-reference styles:  
Example 2.7.3, Example 3; 2.7.3, 3; Related Rates
- Configurable automatic numbering:  
Example 134, Example 2.83, Example 2.7.13, Example 2.7.3.5

# Vocabulary Features

- First priority: author-friendly
- Very few abbreviations: **p**, **ol**, **li**, **c**, **q**, **m**
- Lower-case, few hyphens:  
**theorem**, **proof**, **subsection**, **hint**, **latex-image**
- Limited attributes, sensible defaults
- Robust and unified cross-references with **xref**
- Variety of cross-reference styles:  
Example 2.7.3, Example 3; 2.7.3, 3; Related Rates
- Configurable automatic numbering:  
Example 134, Example 2.83, Example 2.7.13, Example 2.7.3.5
- RELAX-NG grammar (as PreTeXt literate program)
- Schematron rules for exceptions, deprecations

## PreTeXt Example

```
<theorem xml:id="power-rule">
  <title>Power Rule</title>
  <idx>power rule</idx>

  <statement>
    <p>The derivative of <m> $f(x)=x^n$ </m>
    is <m> $f'(x)=nx^{n-1}$ </m>.</p>
  </statement>

  <proof>
    <p>Apply induction to the product
      <me> $f(x)=x^n=x\cdot x^{n-1}$ </me>
      using <xref ref="product-rule"/>.</p>
  </proof>
</theorem>
```

**Theorem 4.4** (Power Rule). *The derivative of  $f(x) = x^n$  is  $f'(x) = nx^{n-1}$ .*

*Proof.* Apply induction to the product

$$f(x) = x^n = x \cdot x^{n-1}$$

using [Theorem 4.1](#).



**Theorem 4.4** (Power Rule). *The derivative of  $f(x) = x^n$  is  $f'(x) = nx^{n-1}$ .*

*Proof.* Apply induction to the product

$$f(x) = x^n = x \cdot x^{n-1}$$

using [Theorem 4.1](#). □

≡ Contents	Index	< Prev	^ Up	Next >
Front Matter	<div><p><b>Theorem 4.4 Power Rule.</b> <i>The derivative of <math>f(x) = x^n</math> is <math>f'(x) = nx^{n-1}</math>.</i></p><p><i>Proof.</i></p><div><p>Apply induction to the product</p><math display="block">f(x) = x^n = x \cdot x^{n-1}</math><p>using <a href="#">Theorem 4.1</a>.</p></div></div>			
1 Introduction				
2 The Fundamental Theorem				
3 Computing Integrals with Sage (f)				
4 An Interesting Corollary				
5 Some Facts and Figures				
6 Some Advanced Ideas	<div><p><b>Corollary 4.5.</b> <i>Suppose <math>f(x)</math> is a continuous function. Then</i></p></div>			

# Separate Content from Presentation (Easier Said Than Done)

- Our authors know  $\text{\LaTeX}$  (escape characters!)
  - Emphasis is `\emph{semantic}`: *semantic*
  - Mathematics is `$x$`:  $x$
  - But `\textit` is a font change: *italics*

# Separate Content from Presentation (Easier Said Than Done)

- Our authors know  $\text{\LaTeX}$  (escape characters!)
  - Emphasis is `\emph{semantic}`: *semantic*
  - Mathematics is `$x$`:  $x$
  - But `\textit` is a font change: *italics*
- Italics (Chicago Manual of Style, 15th edition)
  - Emphasis (7.49)
  - Foreign words (7.51)
  - Key terms (7.57)
  - Titles of books (16.16)
  - Math, single-letter variables (14.9)
  - Scientific names (8.127ff.)
  - Ambiguous ranges, i.e. *ff.* (17.131)
  - Initial 3 letters of enzyme names (8.141)
  - Ships, including spacecraft (8.124)
- `<emphasis effect="bold">` (OpenStax CNXML)



# Design Challenges

- Author's expectations for whitespace in mixed-content (e.g., a blank line in  $\text{\LaTeX}$  is a new paragraph)
- Whitespace in mathematics, computer code
- Display equations, now part of a paragraph
- Lists, now part of a paragraph
- Concessions to layout control
  - Image size controlled by `@width="80%"`
  - No heights, limited use of `@aspect="16:9"`
  - **sidebyside** to go horizontal, subfigures
- Static representation of interactive content (e.g. YouTube)
  - Screenshot (scripted via Javascript "pageres" )
  - QR code points to dedicated standalone HTML page



# Division of Labor

## Author, Publisher, Instructor, and Reader/Student

- **Author** creates XML source: structure and words
- **Publisher** has processing switches:  
numbering depth, knows, visibility of solutions
- **Instructor** hosts course: exercise selection,  
online homework, personal and group annotations
- **Reader/Student**: all of the above for their benefit,  
imagine the textbook as the organizing “course portal”
- Documentation: Author’s Guide, Publisher’s Guide



# What's Next?

- Improve principal output formats beyond  $\text{\LaTeX}$ , HTML
- Frameworks for cohesive styles (CSS,  $\text{\LaTeX}$ )
- Bibliographies via Citation Style Language  
(any XSLT 1.0 solutions out there???)
- Better support for online homework beyond WeBWork  
(MyOpenMath, LMS integration)
- Support more STEM disciplines  
(science, technology, engineering and mathematics)
- Expand support for pre-university courses  
(e.g. 2-page spreads)
- Textbook Research: automated data collection

Thank-you for your attention!

`buzzard.pugetsound.edu/talks.html`

PreTeXt: `pretextbook.org`

Community: Google Groups, GitHub (Join us!)

Developers: RAB, David Farmer, Alex Jordan

Partial support for this work was provided by the National Science Foundation's Improving Undergraduate STEM Education (IUSE) program under Award No. 1626455. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.