

Markdown and Pandoc

Peter Binkley

17 June 2016

This demo

... is on github: [code4lib/yeg/markdown-workshop](https://github.com/code4lib/yeg/markdown-workshop)

(Clone it or download the zip)

Part 1: Markdown

Spirit of Markdown

- John Gruber, 2004, with collaboration of Aaron Swartz
- inspired by ascii email format
- markup does not interfere with the readability of the text
- comprehensible without documentation
- [sustainable](#)

Tools

- edit: any text editor; [Atom](#) or [Sublime Text](#) are examples of Markdown-aware editor
- rendering: [Pandoc](#):

```
pandoc -s test.md -o test.html
```

Ingredients

- punctuation marks: - = > * + - \ [] () _ ' !
- blocks, separated by blank lines
- indents
- John Gruber's [Syntax Guide](#)

Basics

- Paragraphs: separated by blank lines
- **italics**: *italics*
- ****bold****: **bold**

- bullet 1
- bullet 2

Heading

Links

Link

[Anchor](http://www.example.com): [Anchor](http://www.example.com)

Image

![Alt text](https://upload.wikimedia.org/wikipedia/commons/0/00/Escribano.jpg):



Indents

Blockquote: use >

Block quote this is a block quote this is a block quote Block quote
this is a block quote this is a block quote Block quote this is a block
quote this is a block quote

Code block: use tab (or four spaces)

```
Code this is code
And this is code
So is this
```

Advanced: Tables

- not part of classic Markdown
- Pandoc supports [four syntaxes](#)

Simple Tables

Right	Left	Center	Default
12	12	12	12
123	123	123	123

Table: Demonstration of simple table syntax.

Table 1: Demonstration of simple table syntax.

Right	Left	Center	Default
12	12	12	12
123	123	123	123

Multiline Tables

Centered Header	Default Aligned	Right Aligned	Left Aligned
First	row	12.0	Example of a row that spans multiple lines.
Second	row	5.0	Here's another one. Note the blank line between rows.

Table: Here's the caption. It, too, may span multiple lines.

(rendered)

Table 2: Here's the caption. It, too, may span multiple lines.

Centered Header	Default Aligned	Right Aligned	Left Aligned
First	row	12.0	Example of a row that spans multiple lines.

Centered Header	Default Aligned		Right Aligned	Left Aligned
	Second	row		
			5.0	Here's another one. Note the blank line between rows.

Metadata: YAML Block

```
---
title: Markdown and Pandoc
css: css/myslides.css
---
```

Flavours

- no one standard, many flavours
- in the last few years, an effort toward standardization; IETF draft for text/markdown media type
- extensions: MultiMarkdown etc.
- notably: GitHub

My learning project

[jekyll-book](#)

A platform for publishing books using jekyll, with html pages and pdf (etc.) downloads

Part 2: Pandoc

- open-source Haskell application and library for converting markup
- begun 2006 by John MacFarlane (philosopher at UC Berkeley)
- handles “standard” Markdown and provides many extensions

Formats

Reads (full): Markdown, CommonMark, PHP Markdown Extra, GitHub-Flavored Markdown,

Reads (partial): Textile, reStructuredText, HTML, LaTeX, MediaWiki markup, TWiki markup, Haddock markup, OPML, Emacs Org mode, DocBook, txt2tags, EPUB, ODT and Word docx

Writes: plain text, PDF (if it has LaTeX support), Markdown, CommonMark, PHP Markdown Extra, GitHub-Flavored Markdown, reStructuredText, XHTML, HTML5, LaTeX (including beamer slide shows), ConTeXt, RTF, OPML, DocBook, OpenDocument, ODT, Word docx, GNU Texinfo, MediaWiki markup, DokuWiki markup, Haddock markup, EPUB (v2 or v3), FictionBook2, Textile, groff man pages, Emacs Org mode, AsciiDoc, InDesign ICML, TEI Simple, and Slidy, Slideous, DZSlides, reveal.js or S5 HTML slide shows

Generating a PDF

```
pandoc -s slides.md -o slides/slides.pdf
```

[PDF of these slides](#)

Embedding Markup

HTML tags for the HTML god, LaTeX markup unto the PDF generator

```
<p style="text-align: right">
Some right-aligned text in embedded HTML.
</p>
```

```
\begin{flushright}
Some right-aligned text in embedded LaTeX.
\end{flushright}
```

Sample: [HTML](#) ~ [PDF](#)

Filters

Filters can manipulate Pandoc's "abstract syntax tree" (AST) between the parsing of the input and the rendering of the output.

Inspect AST: `pandoc -t native ast.md`

Some **italic** text.

```
[Para [Str "Some",Space,Emph [Str "italic"],Space,Str "text."]]
```

Example Filter in Ruby

- use [pandoc-filter](#) gem
- install: `gem install pandoc-filter`

```
PandocFilter.filter do |type, value, format, meta|
  if type == 'Emph'
    PandocElement.Strong(value)
  end
end
```

- Output: [unfiltered](#) ~ [filtered](#)

Changing the Template

Dump the default template:

```
pandoc -D latex > template.latex
```

```
$if(dedication)$
\begin{center}
$dedication$
\end{center}
$endif$
```

```
pandoc --template=custom.latex custom.md -o slides/custom.pdf
```

Output: [PDF](#)

Special Markup in HTML and PDF

```
<div class="box" markdown="1">
```

```
Brown, John, 1852-1900
\ \ \ The management of institutions,
\ \ \ ed. by Thomas Jones, and rev. by
\ \ \ Henry Hill.....NY.,
\ \ \ Brown - 1938
\ \ \ \ \ \ 440 p.
```

```
</div>
```

```
./preprocess.rb catalog.md | pandoc -o slides/preprocessed-catalog.pdf
```

Output: [HTML](#) ~ [PDF](#) ~ [preprocessed PDF](#)

ePub

An example of an [ePub](#) generated by Pandoc.

Bibliography

Pandoc provides the citations extension and the pandoc-citeproc filter to generate citations and a bibliography. ([Manual](#))

It uses [CSL styles](#)

Sample: [Markdown](#) ~ [BibTex](#) ~ [HTML](#) ~ [PDF](#)

Try another CSL: uncomment `#cs1: apa.cs1` in YAML

Scholarly Markdown?

- an initiative from the sciences (originally Markdown for Science)
- adds better support for inline math, figures, etc.
- implemented in a fork of Pandoc
- See scholarlymarkdown.com (Tim T.Y. Lin)
- [Syntax](#)