Efficient Scientific Research
with (Multi)Markdown

Output 2014
Dipl.-Inf. Thomas Kühn
SO... YOU THINK YOU HAVE WHAT IT TAKES TO BE A SCIENTIST...?!?!
A Small Survey

• Q: What tools do you all use for your daily writing?
• 65 Answers named 34 different Tools
A Small Survey

- Q: *What tools do you all use for your daily writing?*
- 65 Answers named 34 different Tools
A Small Survey

Q: What tools do you all use for your daily writing?

- 65 Answers named 34 different Tools

- Word Processor: 15, 8%
- Latex Environment: 56, 30%
- Text Editor: 114, 62%

Efficient Scientific Research
Survey Results

- Huge amount of tools in use
- Each participant used ~2,6 tools on average
- At least one text editor or a LaTeX environment

Consequences

- Unmanaged redundancy of knowledge between different document formats
- Leads to *Copy-And-Paste* behavior
- Writing performance is different among different tools
- LaTeX environments require the use of latex ideoms

**What is a better strategy for efficient writing?**
Requirements

- Focus on content rather than ideoms
- Concise ways to structure text
- Direct support for citations, figures, tables
- Transformation to arbitrary formats
- **One content source many output formats**
Markdown

- Focus a simple syntax to both *read* and *write*
- Developed by *John Gruber* to write for the web (HTML)

# This is another H1

## This is another H2

*This is an italic text*

_This is a text with an underscore_

**This is a bold text**

The following is a Quotation:

> Lorem ipsum dolor sit amet,

![This is an Image](rosi.png "This is the alt text")

## Lists

* Item of a bullet list
* Another Item

1. An Item of an ordered list
2. Yet another Item

## Highlights

This is a code block:

```
$pandoc --help
```

And this a link to [*Pandoc*](http://johnmacfarlane.net/pandoc/).

1) [http://daringfireball.net/projects/markdown/syntax](http://daringfireball.net/projects/markdown/syntax)
(Multi)Markdown\(^2\)

- Markdown extension developed by *Fletcher T. Penney* to support research activities
- Important features are: *Footnotes, Tables, Citations, ...*

---

# Citations and Footnotes

This is a text to cite [@author], whereas the following Noun[^I am a footnote] carries a footnote.

---

# Special Features

This ~~is striked out text.~~

H\textasciitilde{}2\textasciitilde{}0 is a liquid. 2^{10^\text{

\$\forall x \in X, \exists y \leq \epsilon}$

---

### Tables

<table>
<thead>
<tr>
<th>Fruit</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bananas</td>
<td>$1.34</td>
</tr>
<tr>
<td>Oranges</td>
<td>$2.10</td>
</tr>
</tbody>
</table>

[^I am a footnote]: http://fletcherpenney.net/multimarkdown/
(Multi)Markdown

- Markdown extension developed by Fletcher T. Penney to support research activities
- Important features are: Footnotes, Tables, Citations, ...

\section{Citations and Footnotes}
This is a text to cite \cite{author}, whereas the following Noun\footnote{I am a footnote} carries a footnote.

\section{Special Features}
This \textbackslash out is striked out text. \textbackslash superscript{2} \textbackslash in \textbackslash X, \textbackslash exists \textbackslash y \textbackslash leq \textbackslash \epsilon \textbackslash n \textbackslash \epsilon

\section{Tables}
\begin{longtable}[c]{@{}rlcl@{}}
\toprule
Right & Left & Center & Default \\
\midrule
12 & 12 & 12 & 12  \\
123 & 123 & 123 & 123  \\
\bottomrule
\end{longtable}
\begin{longtable}[c]{@{}ll@{}}
\toprule
\begin{minipage}[b]{0.21\textwidth}
Fruit
\end{minipage} & \[…\]
\bottomrule
\end{longtable}
Pandoc

- Command line tool written in Haskell
- Available for **all** platforms
- Many extensions already included
- Direct transformation from
  - (multi)markdown, reStructuredText,
  - HTML, LaTeX, ...
- To
  - XHTML, HTML 5,
  - LaTeX (including beamer slide shows),
  - ConTeXt, ODT, Word docx,
  - EPUB (v2 or v3), ...
- Supports use of customized templates

```bash
$ pandoc -S test.md -o test.tex
```

3) http://johnmacfarlane.net/pandoc/
Efficient Scientific Research
**Templates**

- Pandoc's templates contain variables, conditionals, and for-loops
- `$body` is a special variable for the actual markdown content
- Large number of predefined templates

**Using Templates**

- Add `yaml_metadata_block` extension
- Variable definition at the beginning of the document

```shell
$ pandoc -s -S "task.md"
-f markdown+yaml_metadata_block
-t latex --template="ttask.tex"
```
Missing Features

- Include files into documents
- Conditional creation of content
- Embed source code in documents

PanRuby

- Ruby-based command line tool
- Employing ERB as template engine

```
<% unless variant? == :beamer %>
## Overview
<% end %>
```

```
<%=
  load("sections/sec2.md")
<%=
  load("sections/sec3.md")
<%=
  load("sections/sec4.md")
<%=
  load("sections/sec5.md")
<%=
  load("sections/sec6.md")
<%=
  load("sections/sec7.md")
<%=
  load("sections/ack.md")
%
```

```
$ ruby build.rb latex paper
```

Efficient Scientific Research

Efficient Writing
Use Case 1 - Presentations

- Make your notes in (Multi)Markdown and convert them directly to slides

```bash
$ ruby build.rb beamer presentation
  ../trosi_presentation.tex
$ mlatex presentation
```

Please note, that `mlatex` is an alias for `pdflatex` and `bibtex`, whereas the former is called multiple times.
Conclusion

- Expressibility of (Mulit)Markdown sufficient for scientific writing
- Files written in Markdown are easy to read and write
- Tool chain is platform independent
- Everyone can use their favorite Text Editor
  - Get fast with it
  - Write macros and helper functions
- Export your documentation to arbitrary formats
- Design your layouts and templates once

**Focus on content rather than LaTeX ideoms**
Automated Organization

- Automated BibTex lookup for stored papers
- Automated BibTex lookup for specific publications from the web
- Automated filtering of large BibTex files
- Automated download of papers referenced by a BibTex file
"I spend a lot of time on this task. I should write a program automating it!"

Theory:
- Writing code
- Work on original task
- Automation takes over
- Free time

Reality:
- Writing code
- Debugging
- Ongoing development
- Rethinking
- No time for original task anymore

Automation (By Randell Munroe from xkcd.com)