Using \[M\] inside \TeX\ Documents

TUG@Bacho\TeX\ 2017

https://github.com/witiko/markdown

Vít Novotný
Contents

1. Introduction
   1.1 The Case for Lightweight Markup
   1.2 Existing Solutions

2. The markdown.tex Package
   2.1 Building a Parser
   2.2 Quick Overview of the Package

3. Conclusion

4. Q&A

5. Bibliography
Section 1

Introduction
The Case for Lightweight Markup

\textit{\LaTeX} as a Content Creation Language

1. High Markup to Text Ratio
   - \textit{The \LaTeX\book} (Knuth, 1986) is 22\% markup (plain \LaTeX).
   - \textit{Think Java} (Downey et al., 2016) is 21\% markup (\LaTeX).

2. Zero Sandboxing Support
   - The document you are typesetting may not compile.
     \texttt{innocent\_looking\_underscores.tex}
   - The document you are typesetting may halt.
     \texttt{\def\whiletrue{\whiletrue} \whiletrue}
   - The document you are typesetting may access the system shell.
     \texttt{\immediate\write18{sudo \texttt{rm -rf /}}}

3. Steep Learning Curve
The Case for Lightweight Markup

Comparison of \LaTeX and Markdown

\section{This is a level one heading}
This is a text paragraph with \textit{emphasis}.
\begin{quotation}
This paragraph will show as a quote.\end{quotation}
\begin{verbatim}
This is a source code example.
\end{verbatim}
\begin{itemize}
\item First item with \textbf{strong emphasis}
\item Second item with a link\footnote{See \url{http://link.com} (Title)}
\end{itemize}
\begin{enumerate}
\item First item with \verb inline code\.
\item Second item with an \includegraphics{image.png}
\end{enumerate}
The Case for Lightweight Markup
Comparison of $\LaTeX$ and Markdown

# This is a level one heading
This is a text paragraph with _emphasis_.
> This paragraph will show as a quote.

This is is a source code example.

* First item with **strong emphasis**
* Second item with a [link](http://link.com/ "Title")

1. First item with `inline code`.
2. Second item with an ![image](image.png "Title")
The Case for Lightweight Markup

*Markdown as a Content Creation Language*

1. **Minimal Markup to Text Ratio**
   - Recall: Knuth (1986) and Downey et al. (2016) are ~22% markup.
   - *Efficient R programming* (Gillespie et al., 2016) is 5.5% markup.
   - *R for Data Science* (Grolemund et al., 2016) is 3.8% markup.

2. **Either Sandboxing Support ...**
   - A Markdown document converted to \TeX{} will always compile.
   - The document may neither halt nor access the shell.

3. **... or Hybrid Markup Support**
   - Structurally simple sections can use pure Markdown, complex sections may combine Markdown and the host markup.

4. **Mild Learning Curve**
**Existing Solutions**

*The Swiss Army Knife of Pandoc*

*If you need to convert files from one markup format into another, Pandoc is your swiss-army knife.*

— MacFarlane (2016b), emphasis mine

- A multi-target publishing software.
- Supports tens of markup languages (Markdown, \( \LaTeX \), HTML, XML Docbook) and output formats (ODF, OOXML, PDF).
- The use of Pandoc for the preparation of \( \LaTeX \) documents has been described in TUGBoat by Dominici (2014).
Existing Solutions
Why Is Pandoc Not Ideal?

1. Difficult to Change Output Markup

# Heading {#link}
This is [a link](#link).

\hypertarget{link}{\section{Heading}\label{link}}
This is \protect\hyperlink{link}{a link}.

2. Not a Part of \TeX Distributions
   - Markdown documents cannot be directly edited at collaborative \TeX platforms such as Share\TeX or Overleaf.
Existing Solutions

Why Is Pandoc Not Ideal?

3. Half-hybrid, Half-sandboxed

- The input is heuristically parsed and sanitized:

This \{will\} $2^n$ \begin\{get\} s~nitized and \this\{will\} not \begin\{equation\}$2^n$\end\{equation\} $2^n$.

\[ \text\{will\} 2^n \begin\{get\} s~nitized and \this\{will\} not \begin\{equation\}2^n\end\{equation\} $2^n$. \]

- Malicious input such as

\def\shell{18} \immediate\write\shell\{sudo rm -rf /\}

is left alone by Pandoc.
Section 2

The markdown.tex Package
Building a Parser

Is \TeX\ Up to the Task?

There exist formal language parsers written solely in \TeX. These parsers recognize regular (\TeX3 Project, 2016) and context-free LL(1) languages (Carlisle, 2000). Markdown is not context-free:

```
```
There is a literal backtick (``) here.
```

and a parser needs to be able to backtrack over the entire input:

[Leci n'est pas un link](http://link.com/ "Link"

Implementing such a parser in \TeX\ is possible, but generally a bad idea due to the lack of efficient data structures.
Can We Use Lua Instead of \TeX?

Lua is a powerful, efficient, lightweight, embeddable scripting language. It supports procedural programming, object-oriented programming, functional programming, data-driven programming, and data description.

— Lua Team (2016)

\LaTeX{} is an extended version of pdf\LaTeX{} using Lua as an embedded scripting language.

— \LaTeX{} Team (2016)
Building a Parser
Can We Use Lua Instead of \TeX? 

- With Lua\TeX{}, we can directly execute Lua code:

```
1 + 2 = \texttt{directlua}\{ \texttt{tex.sprint(1 + 2)} \}
```

- With pdf\TeX{} and other modern \TeX{} engines, we can spawn a shell and execute the Lua code in a separate process:

```
1 + 2 = \texttt{newwrite}\texttt{script}
\texttt{immediate}\texttt{openout}\texttt{script}=\texttt{script.lua}
\texttt{immediate}\texttt{write}\texttt{script}\{ \texttt{print(1 + 2)} \}\%
\texttt{immediate}\texttt{closeout}\texttt{script}
\texttt{immediate}\texttt{write18}\{ \texttt{texlua script.lua > output.tex} \}\%
\texttt{input} output.tex
```
Building a Parser

The Lunamark Library

- Lunamark (MacFarlane, 2016a) is a Markdown parser in Lua.
- The language is specified using a Parsing Expression Grammar (PEG) via the L Peg C library (and a bit of cheating).
- The dependencies of Lunamark were all either compiled into LuaTeX (L Peg, Slunicode), or unnecessary (Cosmo, Alt-getopt).
- The library has been released under the Expat (MIT) License.
Building a Parser
A High-Level Overview

A modified version of Lunamark transforms an input Markdown document into a plain \TeX document that encodes its structure:

\begin{verbatim}
# Heading
This is [a link](#link).
\end{verbatim}

\begin{verbatim}
\markdownRendererHeadingOne{Heading}
This is \markdownRendererLink{a link}{#link}{#link}{}.
\end{verbatim}

A plain \TeX macro package defines the \markdownRenderer... macros and typesets the above document.
Quick Overview of the Package

A Block Diagram of the Package

User Code

Con\TeX\ Interface

Plain\ TeX\ Interface

Lua Interface
Quick Overview of the Package

The Lua Interface

```lua
#/usr/bin/env texlua
local kpse = require"kpse"
kpse.set_program_name"kpsewhich"
local markdown = require"markdown"
local convert = markdown.new({ something = true })
local input = "Some _Markdown_ text."
print(convert(input))
```
Quick Overview of the Package
The Plain \TeX\ Interface

\input markdown 
\def\markdownOptionSomething{true}
\def\markdownRendererSomethingElse#1#2#3{\foo{#1}}
\markdownInput{bar.md}
\markdownBegin
Some _Markdown_ text.
\markdownEnd
\bye
Quick Overview of the Package

The Con\TeX{}t Interface

\usemodule[t][markdown]
\def\markdownOptionSomething{true}
\def\markdownRendererSomethingElse#1#2#3\{\foo\{#1\}\}
\starttext
\markdownInput\{bar.md\}
\startmarkdown
Some _Markdown_ text.
\stopmarkdown
\stoptext
Quick Overview of the Package

The \TeX\!2e Interface

\documentclass{article}
\usepackage[something]{markdown}
\markdownSetup{renderers = {somethingElse = {\foo{#1}}}}
\begin{document}
\markdownInput{bar.md}
\begin{markdown}
Some _Markdown_ text.
\end{markdown}
\end{document}
Quick Overview of the Package

*The Sandbox and Hybrid Modes*

```latex
\documentclass{article}
\usepackage{markdown}
\begin{document}
\begin{markdown}
Foo bar \textbackslash TeX\{}
$2^n$.
\end{markdown}
\begin{markdown*}{hybrid}
Foo bar \textbackslash TeX\{}
$2^n$.
\end{markdown*}
\end{document}
```

Foo bar \textbackslash TeX\{} $2^n$. Foo bar \textbackslash TeX\{} $2^n$.  

21/44
Quick Overview of the Package

Mapping Markdown Tokens to TeX Macros

\documentclass{article}
\usepackage{markdown}
\markdownSetup{renderers = {
  link = {#1\footnote{See \url{#3} (#4)}},
}}
\begin{document}
\begin{markdown}
  Foo [bar](http://link.com "Link").
\end{markdown}
\end{document}

Foo bar\textsuperscript{1}.

\textsuperscript{1}See http://link.com (Link)
Quick Overview of the Package

Syntax Extensions

- Some syntax extensions were already supported by Lunamark:
  - HTML,
  - footnotes,
  - definition lists,

- New syntax extensions were added as a part of the project:
  - citations,
  - fenced code blocks,
  - IA Writer content blocks.
Quick Overview of the Package

Syntax Extensions – \markdownSetup{html}

HTML <b>tags</b> such as &lt;b&gt; are recognized
<!-- and comments are just ignored-->

HTML tags such as <b> are recognized.
Quick Overview of the Package

Syntax Extensions – `{markdownSetup}{footnotes}`

Here is a footnote reference,[^1] and another.[^long]

[^1]: Here is the footnote.

[^long]: Here’s one with multiple blocks.

Subsequent paragraphs are indented to show that they belong to the footnote.

Here is a footnote reference,^[2] and another.^[3]

^[2]: Here is the footnote.

^[3]: Here’s one with multiple paragraphs.

Subsequent paragraphs are indented to show that they belong to the footnote.
Quick Overview of the Package

Syntax Extensions – `markdownSetup{definitionLists}`

Term 1
:  Definition

Term 2
:  Definition with multiple paragraphs

Term 1  Definition 1

Term 2  Definition
    with multiple paragraphs
Quick Overview of the Package

Syntax Extensions – \markdownSetup\{citations\}

Here is a parenthetical citation [@knuth86] and a string of several [see @knuth86, pp. 33-35; also @gruber04, chap. 1].

Here is a text citation @knuth86 and a string of several @knuth86 [pp. 33-35; @gruber04, chap. 1].

Here is a parenthetical citation (Knuth, 1986) and a string of several (see Knuth, 1986, pp. 33-35; also Gruber, 2004, chap. 1). Here is a text citation Knuth (1986) and a string of several Knuth (1986, pp. 33-35) and Gruber (2004, chap. 1).
Quick Overview of the Package

Syntax Extensions – `markdownSetup{fencedCode}`

```js
if (a > b)
    return c + 4;
else
    return d + 5;
```

```js
if (a > b)
    return c + 4;
else
    return d + 5;
```
Quick Overview of the Package

Syntax Extensions – \markdownSetup{contentBlocks} /

/Flowchart.png "Engineering Flowchart"

Figure: Engineering Flowchart
Quick Overview of the Package

Syntax Extensions – \markdownSetup{contentBlocks} II

/Scientists.csv (Great Minds of the 19th Century)

<table>
<thead>
<tr>
<th>name</th>
<th>surname</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albert</td>
<td>Einstein</td>
<td>133</td>
</tr>
<tr>
<td>Marie</td>
<td>Curie</td>
<td>145</td>
</tr>
<tr>
<td>Thomas</td>
<td>Edison</td>
<td>165</td>
</tr>
</tbody>
</table>

Table: Great Minds of the 19th Century
Quick Overview of the Package

Syntax Extensions – \markdownSetup{contentBlocks} III

/chapters/01.txt
/chapters/02.txt

Chapter 1
This is the first chapter.

Chapter 2
And this is the second chapter.
Quick Overview of the Package

Syntax Extensions – `markdownSetup{contentBlocks}` IV

https://tug.org/tugboat/noword.jpg
(The Communications of the \TeX{} Users Group)

Figure: The Communications of the \TeX{} Users Group

(This actually does not work out-of-box at the moment.)
Section 3

Conclusion
Conclusion

The Missing Pieces of the Puzzle

The `markdown.tex` package

- enables the use of Markdown in environments where tools from outside TeX distributions are unavailable,
- gives the authors full control over how individual Markdown elements are rendered and how much access to TeX markup the Markdown documents have,
- exposes Lua, plain TeX, ATEX, and ConTeXt interfaces.
- was released under the LaTeX Project Public License (LPPL) 1.3 on the Comprehensive TeX Archive Network (CTAN) and on GitHub ([https://github.com/witiko/markdown](https://github.com/witiko/markdown)).
Conclusion
The Missing Pieces of the Puzzle

- The syntax extensions were backported to Lunamark and merged by MacFarlane, resulting in a new minor version release of the library (0.5.0). (Novotný, 2016a)

- The package was featured on the twitter profile and the blog of Overleaf – a major online service for preparing \LaTeX{} documents – along with original example documents. (Lim, 2017)

- \TeX{} comments in hybrid code behave in an unexpected way.

- Logging should be improved, so that programs such as \texttt{latexmk} can automatically remove cache files that are no longer needed.
Section 4

Q&A
Section 5

Bibliography
Bibliography I

\TeX\ PROJECT, 2016. \textit{The l3regex package: regular expressions in \TeX} [online] [visited on 2016-11-08]. Available from: http://mirrors.ctan.org/macros/latex/contrib/l3experimental/l3regex.pdf.

Bibliography II


Bibliography III


Bibliography IV


Bibliography V


LUAT\TeX{} TEAM, 2016. *Lua\TeX{}: Welcome* [online] [visited on 2016-08-15]. Available from: http://luatex.org/.
Bibliography VI


