Pandoc for PDF How-To

How-to generate nice PDF documentation from Markdown

Alexey Gumirov

17 December 2018

Contents

Pandoc for PDF How-To

How-To, templates and commands to produce PDF documents from MarkDown files.

How-to for docs preparation

Tools

- **pandoc**

    - template: Original web-site eisvogel.latex[1].
    - The same template file in this project repository in our GitLab[2]

- **texlive**

- **convert**

    - converts and formats images.
    - it is used here for the change of DPI of the images and convert to PNG.
    - **convert** is the utility which is part of the **ImageMagick** package.

I did not install **convert** tool, it seems like it is installed by default in Ubuntu or comes with **texlive**. To avoid possible issues with **pdflatex** engine I did full installation of **texlive** packet.

In Debian family (with **apt**):

```
1  sudo apt-get update
2  sudo apt-get install pandoc
3  sudo apt-get install texlive-full
4  sudo apt-get install imagemagick
```

Because of the size of the **texlive** packet (~4,5GB) I do generation of the PDF locally.

Instructions and commands

YAML Block for LaTex template

This YAML block in the beginning of the MarkDown file defines parameters used by the Pandoc engine and relevant LaTex template (in my case - **Eisvogel**). This particular example below instructs Pandoc to produce PDF file with the Cover page (*titlepage: **true***) and change color of the line on the cover page.

---

[1]https://github.com/Wandmalfarbe/pandoc-latex-template
[2]pandoc/templates/eisvogel.latex

Another important parameter is **logo** - it defines path to file with the logo you want to put on the cover page.

```
1   title: "Pandoc for PDF How-To"
2   author: "Alexey Gumirov"
3   date: "17 December 2018"
4   subtitle: "How-to generate nice PDF documentation from Markdown"
5   titlepage: true
6   titlepage-color: "FFFFFF"
7   titlepage-text-color: "000000"
8   titlepage-rule-color: "CCCCCC"
9   titlepage-rule-height: 4
10  toc-own-page: true
11  logo: "files/logo.png"
12  logo-width: 100
13  links-as-notes: true
14  lof: false
15  lot: false
```

Parameter **links-as-notes** enables putting of the URL links in the footnotes of the page.

Because GitHub does not support YAML header in the main file, I set it up in the separate file in the root folder of the project. I call it **_yaml-block.md**.

Images preparation

In my setup I print with 300 DPI (this produces high resolution PDF). Therefore all images must be 300 DPI. If you have images with different DPI (especially GIF files), then use the following commands:

To re-sample image to 300 DPI:

```
1   convert $SOURCE_IMG_FILE -units PixelsPerInch \
2       -resample 300 $TARGET_IMG_FILE.png
```

After rasampling image has to be brought to the proper size. Command resizes picture to be 1700 pixels horizontally and sets DPI meta-data to 300.

```
1   convert $SOURCE_IMG_FILE -units PixelsPerInch \
2       -resize 1700x -density 300 $TARGET_IMG_FILE.png
```

But if you are not afraid, then all can be done in one command:

```
1   convert $SOURCE_IMG_FILE  -set units PixelsPerInch \
2       -resample 300 -resize 1700x -density 300 $TARGET_IMG_FILE.png
```

It is important to mention, that order of options does matter. The instruction above makes steps in the following order:

1. `-set units PixelsPerInch`: Sets density units in Pixels per Inch instead of default `PixelsperCantimeter`.
2. `-resample 300`: Changes resolution of the image from its current DPI (PPI) to 300 DPI (PPI). It is not just change of meta-data, this parameter makes **convert** to re-process image.
3. `-resize 1700x`: Resizes picture to the following dimentions: `width` = 1700 `pixels`, `height` = `auto`.
4. `-density 300`: This parameter sets DPI meta-data in the target picture to 300 DPI (PPI)

Pandoc command

```
1  pandoc -s -S -o $DEST.pdf --template eisvogel \
2      --toc --dpi=300 -V lang=en-US _yaml-block.md $SOURCE.md
```

Parameters of the **pandoc** command mean following:

- `-s`: Standalone document.

- `-S`: `--smart`

  - Produce typographically correct output, converting straight quotes to curly quotes, — to em-dashes, – to en-dashes, and … to ellipses. Nonbreaking spaces are inserted after certain abbreviations, such as "Mr." (Note: This option is selected automatically when the output format is latex or context, unless `--no-tex-ligatures` is used. It has no effect for latex input.)

- `--template FILE`: Use `FILE` as a custom template for the generated document. Implies `--standalone`.

- `--toc`: `--table-of-contents`

  - Include an automatically generated table of contents (or, in the case of latex, context, docx, and rst, an instruction to create one) in the output document. This option has no effect on man, docbook, docbook5, slidy, slideous, s5, or odt output.

- `--dpi`:

  - Specify the dpi (dots per inch) value for conversion from pixels to inch/centimeters and vice versa. The **default** is **96dpi**. Technically, the correct term would be ppi (pixels per inch).

- `-V KEY[=VAL]`: `--variable=KEY[:VAL]`

  – Set the template variable KEY to the value VAL when rendering the document in standalone
    mode. This is generally only useful when the `--template` option is used to specify a custom
    template, since pandoc automatically sets the variables used in the default templates. If no
    `VAL` is specified, the key will be given the value true.
  – `lang`: one of the `KEY` parameters of `-V` which defines default document language.

Additional useful options of the **pandoc** command are:

- `--listings`: It creates nice presentation of the raw code (like shell code or programming code).
- `--number-section`: Automatically creates enumerated headers. It is used in the Lesson
  Learned - vEPC example below.
- `--default-image-extension`: If you want Pandoc to insert only one type of images, e.g. PNG,
  then you shall add `--default-image-extension png` in the command line.

Important notes about MarkDown file formatting for PDF processing

Unordered Lists and sub-lists indentation

It is stated in the GitHub[3] site that correct indent for the unordered lists is 2 spaces. But with this indent
Pandoc does not identify sub-lists.

Therefore, please use 4 spaces indent for the sub-lists in the unordered lists. Then they will be properly
reflected in the PDF files.

While using of standard tab (4 spaces) indent is not a mistake, some programs (in my case it is MS Visual
Studio Code) can give you a warning. You can just ignore it.

Links

If your Markdown file has to be processed into the PDF, then better not to use format of links, which
open in a separate window, e.g.:

a) Link format that does NOT work: `![Name of the resourse](Link)`.
b) Links that WORKS: `[Name of the resource](Link)`.

The problem is that by the Markdown guidelines using exclamation mark before URL is not appropriate.
But GitLab engine does not give you an error, it just treats such links as links which open in the new tab
or window in the browser. Therefore, to avoid compilation errors in the **pdflatex** engine (which is used
by **pandoc**), please use (b) type of URL formatting, whcih is compliant with Markdown standard.

---

[3]https://github.com/DavidAnson/markdownlint/blob/v0.11.0/doc/Rules.md#md007

Examples

This page example

This page pandoc-2-pdf-how-to.pdf. Generated with the following command (in the project directory):

```
1  pandoc -s -S -o pandoc-2-pdf-how-to.pdf --template eisvogel \
2      --toc --listings --dpi=300 \
3      -V lang=en-US _yaml-block.md README.md
```

The link to _yaml-block.md file is here[4].

---

[4] _yaml-block.md