Converting Word documents into Markdown

This documented shows how to take a word document and convert it into Pandoc flavoured Markdown.

Introduction:

For better or worse there are many examples of technical documentation that is maintained inside MS Word (or Google docs etc).

The problem is that it’s very hard to:

* Maintain the content under version control
* Re-purpose the content for other formats

# Approach

Translate the word document into a markdown document as a once-off migration step

You need to have [Pandoc](https://pandoc.org/) installed

The process is can be pictured as follows:

Figure : Pandoc Workflow

Export the content with the following command

pandoc -f docx -t markdown[[1]](#footnote-1) file.docx -o file.pmd

Try it!

# Images

But where is the imbedded picture?

We need to export images separately and then add a link to the Markdown file by hand. We only need to do this once after exporting the Word content.

In MS Word the easiest way to export all the images is via “Save As” and using a file format of “Web Page (filtered)”. Use a file name of “images.html” an all the images in the document will then be exported to a subdirectory called “images.fld”. (Note that these options may be different depending on the version of MS Word you are using.)

Then manually add image links in your Markdown file. Edit the file “file.pmd” and replace the orphaned image caption with a markdown image link as follows

![Pandoc Workflow](images.fld/image001.png “MS Word files being filtered by Pandoc into Markdown”)

Put the text on a single line with a blank like above and below (and don’t use smart quotes).

Save the file and then review the content in html

pandoc -f markdown -t html file.pmd -o file.html

and open the file.html

If this is done correctly then an HTML version of the original MS Word content is displayed with the embedded image (including alt text), working links and a footnote with link.

1. The ***markdown*** format creates Pandoc flavoured markup. See the Pandoc manual for details. [↑](#footnote-ref-1)