

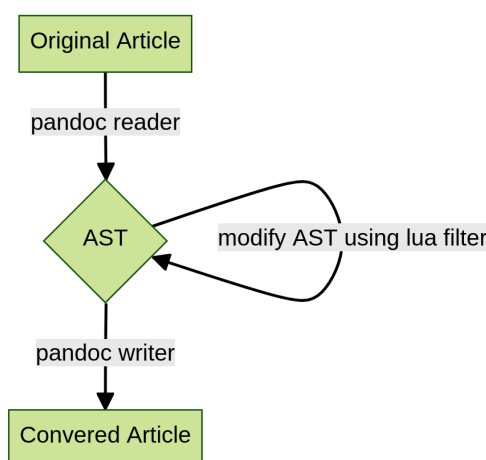
# Understanding pandoc lua filters

by Abhishek Ulayil

**Abstract** pandoc supports intermediate modification of the Abstract Syntax Tree (AST) between the parsing and writing phase using filters. This supplement paper will highlight the use cases of such filters written in Lua Language on LaTeX, markdown and native AST.

## 1 Introduction

To understand pandoc filters, one must look at the document conversion process in general. A document is read/parsed in one format to a common intermediate format. In pandoc it is called an abstract syntax tree (AST). A document writer is designed to read the AST and write the contents in the form of desired article format. For example converting markdown to HTML will involve these processes in pandoc.



**Figure 1:** Conversion Workflow

The filters come into picture when there are certain elements which require customization. For example, suppose you are converting markdown to HTML and you require the page to have automatic numbering for figures or tables. If you think about it in markdown there is no system of numbering figures/tables automatically, to retain 1:1 conversion the HTML output will miss the numbering as well. At this moment you would wish that there was some option to add such a feature, but then given the infinite possibilities of customization it is not feasible to add each and every customization as an option in pandoc or the writer. To tackle this, the idea of filters arrived, wherein you could manipulate the AST partly or fully and modify it before the writer could read the AST. Hence generating an output you desire (MacFarlane and other pandoc authors, 2023).

## 2 Writing lua filters

Continuing the example above lets create a dummy article, where we need to add numbering. One way could be to keep a counter of Images, and add a prefix "Figure X : " to each figure caption. This will serve the purpose of numbering the Images in the end result.

```
![R logo](Rlogo-5.png){width="10%"}
```

```
![penguins](penguins.png){width="15%"}
```

```
pandoc example.md --from markdown --to html5 --output example.html
```

**Figure 2:** pandoc command without lua filter or extensions

Now If we convert the above markdown file to HTML5 using pandoc command 2, we get 3



Figure 3: Vanilla HTML5 output

### 3 Using pandoc lua filters

As we can see, there is no figure numbering done automatically, which is generally the expected result. However if we want to include numbering to it, we would need to write a lua filter. This lua filter will modify the AST and make the changes we desire.

We call a lua filter in the pandoc command [2](#) using the `--lua-filter name_of_filter.lua` option in pandoc.

Now we write a lua filter to manipulate the figures in [4](#)

**TODO : explain lua filter in more detail.**

```
figures = 0
is_fig = 0

function Figure(el)
  local label = ""
  pandoc.walk_block(el, { Image = function(el)
    is_fig = 1
  end })
  if is_fig == 1 then
    figures = figures + 1
    label = "Figure " .. tostring(figures) .. ":"
  end
  local caption = pandoc.utils.stringify(el.caption)
  if not caption then
    caption = label
  else
    caption = label .. " " .. caption
  end
  el.caption.long[1] = caption
  is_fig = 0
  return el
end
```

Figure 4: A lua filter to add figure numbering

## 4 Interpreting changes using lua filters

The filter 4 when included in the pandoc command will generate 6 using the 5 command.

```
pandoc example.md --from markdown --to html5 --output filtered-example.html  
--lua-filter image_numbering_filter.lua
```

Figure 5: pandoc command with lua filter



Figure 1: R logo



Figure 2: penguins

Figure 6: HTML5 output with desired filtering

## 5 Usage of pandoc lua filters in texor package

## 6 Summary

## 7 Supplementary materials

The example conversion files described here are included as supplementary materials in the Rjournal article of texor.

## Bibliography

A. Krewinkel and A. Lucero, pandoc 3.0 Release notes, *pandoc* 2023 URL <https://pandoc.org/releases.html#pandoc-3.0-2023-01-18> [p]

John MacFarlane and other pandoc authors, Pandoc Lua filters, *pandoc documentation* 2023 URL <https://pandoc.org/lua-filters.html> [p1]

Abhishek Ulayil  
Student, Institute of Actuaries of India  
Mumbai, India  
ORCID: 0009-0000-6935-8690