count-para

Michael Cysouw

Adding a paragraph counter to text

[1] **Summary**: Use this Pandoc filter to add numbers to paragraphs of a text. These numbers can be used in scholarly references in an online world of adaptive design where page numbers are not useful anymore.

- Check out this readme in HTML and PDF.
- Refer to paragraphs by using a hash, e.g. (Cysouw 2021: #9), and add this hashtag to the end of a link to be automatically directed to the paragraph, e.g. (Cysouw 2021: #9).

Rationale

[2] Even scientific writing, notwithstanding its very conservative nature, is moving to new forms of publication. One of the ongoing changes is the appearance of a diverse set of **electronic forms of publication**. Articles and books are all (coming) online, and many are also available as html. Also ePub is clearly an avenue to consider. In general, any such new formats will have to be adaptive to different screen sizes.

[3] One of the central problems this poses for academia (besides some technical issues) is that one of the central scholarly traditions breaks down, namely the **reference to a page number** of a work. Currently, there is a tendency to always *also* make a PDF available that looks like a classical printed book with page layout. The main reason for these PDF-versions is that they still provide an authoritative separation of a work into pages (less and less readers are actually printing them).

[4] It is time to move to a different approach to citation. Figures and tables can be numbered, and of course there are numbered sections. However, for a more fine-grained approach I propose to simply number **paragraphs**. A paragraph is a clearly defined part, and it remains the same paragraph on different screen sizes with different wrappings (this does not hold, for example, for line numbers). A good written paragraph has a clear message, so for precise citation this seems like an ideal match (and ranges of paragraphs can of course also be used).

[5] The idea is to simply put a number at the start of each paragraph, hard coded into the text so it cannot be changed (it has to be a **stable identifier**!). Depending on the electronic format, different ways to actually show these identifiers can be used. Being hard coded at the start of a paragraph does not necessarily mean obtrusive!

Practice

[6] There are many ways in practice to add the numbers. I provide here one possibility, a Pandoc Lua filter. Pandoc is a conversion system, mainly provided by John MacFarlane, that allows for a highly flexible conversion between many different output formats. The basic idea is to write text in Pandoc's extension of Markdown, and then the text can be published into different formats. Pandoc has an extension mechanism called 'filters', and the easiest to use variant are 'lua filters'. They basically consist of just one file that provides the extra functionality.

[7] The file `count-para.lua` in this repository is such a file. Used with Pandoc it will count paragraphs, add a number to the front, and provides (currently) nice outputs for HTML and Latex. As an example, this readme-document is provided with paragraph-numbering in HTML and as a PDF made with Latex by the following commands (much of this can be specified in so-called `default` files, which are much easier to handle).

```
pandoc readme.md -t html -o readme.html -L count-para.lua -s
pandoc readme.md -o readme.pdf -L count-para.lua
```

[8] For a more involved example, check out my manuscript about German grammar written in Pandoc Markdown. The HTML version can be directly accessed here.

Referencing

[9] Now we can refer to paragraphs! Of course we can simple write them in our citations, just like page numbers. I propose to use the hash `#` as an indicator of paragraphs, like this (Cysouw 2012: #2.7), to differentiate it from page numbers, like this (Cysouw 2012: 34). This hash is not just a typographic issue, but it is actually the method to link to the paragraph in question. If you add #2.7 to the end of the link to my manuscript you will immediately end up at that paragraph, e.g. click here: (Cysouw 2012: #2.7). By the way, the number is 2.7 because I have restarted numbering at chapters. So this is actually the seventh paragraph of chapter 2. This is an option of the filter by specifying **resetAtChapter: true** in the defaults-file. The term 'chapter' is a slight misnomer, because it simply refers to the highest level of headings in the manuscript.