

## Bookbinding and Wikisource

**Sunday, February 11th, 2018. 12:10PM.**

I have been experimenting with turning Wikisource works into LaTeX-formatted bindable PDFs. My initial idea was to produce quarto or octavo layout sheets (i.e. 8 or 16 book pages to a sheet of paper that's printed on both sides and has the pages layed out in such a way as when the sheet is folded the pages are in the correct order) but now I'm thinking of just using a print-on-demand service (hopefully Padiapress, because they seem pretty brilliant).

Basically, my tool downloads all of a work's pages and subpages (in the main namespace only; it doesn't care about the method of construction of the work) and saves the HTML for these, in order, to a `html/` directory. Then (here's the crux of the thing) it uses Pandoc to create a set of matching TeX files in an adjacent `latex/` directory.

So far, so obvious. But the trouble with this approach of wanting to create a separate source format for a work is that there are changes that one wants to make to the work (either formatting or structural) that can't be made upstream on Wikisource — but we *also* want to be able to bring down updates at any time from Wikisource. That is to say, this is creating a fork of the work in a different format, but it's a fork that needs to be able to be kept up to date.

My current solution to this is to save the HTML and LaTeX files in a Git repository (one per work, e.g.) and have two branches: one containing the raw un-edited HTML and LaTeX, on which the download operation can be re-run at any time; and the other being based off this, being a place to make any edits required, and which can have the first merged into it whenever that's updated. This will sometimes result in merge conflicts, but for the most part (because the upstream changes are generally small typo fixes and the like) will happen without error.

Now I just want to automate all this a little bit more, so a new project can be created (with GitHub repo and all) with a single (albeit slow!) command.

The output ends up something like *The Nether World* by George Gissing.pdf.