Using `tidy_bib`

Paul C. Bauer* Denis Cohen†

2020-06-23

An example

In this example, we use the `tidy_bib()` function to

1. combine two separate `.bib`-files, `references1.bib` and `references2.bib`
2. clean and repair the resulting bib-files (e.g., by removing unwanted fields such as the ISSN, ISBN, DOI, and URL)
3. freely cite works from the combined `.bib`-files in an RMarkdown document which consists of a main text and an appendix
4. create separate `.bib`-files for the main text and the appendix, each containing only those entries which were cited in the respective sections
5. embed the corresponding separate bibliographies for the main text and the appendix

YAML Header

In order for this to work, we must specify some information in the YAML header of our `.Rmd`file:

```
output:
  bookdown::pdf_document2:
    number_sections: no
    toc: false
    pandoc_args: --lua-filter=multiple-bibliographies.lua
    keep_tex: no

header-includes:
  - \usepackage {hyperref}
  - \hypersetup {colorlinks = true, linkcolor = blue, urlcolor = blue}

bibliography_main:
  - partial_bib_1.bib
bibliography_app:
  - partial_bib_2.bib
```

*University of Mannheim, paul.bauer@mzes.uni-mannheim.de
†University of Mannheim, denis.cohen@mzes.uni-mannheim.de

```
link-citations: yes
linkcolor: blue
```

There are two important additions.

1. `pandoc_args: --lua-filter=multiple-bibliographies.lua` calls the lua-filter `multiple-bibliographie` for the creation and inclusion of multiple bibliographies using `pandoc-citeproc`.
2. The definitions of `bibliography_main` and `bibliography_app` specify the suffixes of our partial bibliographies as well as the file names of the corresponding `.bib`files, which will be produced by `tidy_bib()` in the next step.

`tidybib()` code chunk

After the YAML header, make sure to include a (hidden, yet evaluating) R code chunk in your RMarkdown document. Here, we specify the following:

```
tidy_bib(
  rmarkdownfile = "manuscript.Rmd",
  old_bib_file = c("references1.bib", "references2.bib"),
  new_bib_file = "partial_bib.bib",
  by_sections = c("<!--- appendix split -->"),
  repair = TRUE,
  replace_curly_braces = FALSE,
  removeISSN = TRUE,
  removeISBN = TRUE,
  removeDOI = TRUE,
  removeURL = TRUE
)
```

- `rmarkdownfile = "manuscript.Rmd"` specifies that the very same RMarkdown script in which we are writing our paper will be scanned for citations.
- `old_bib_file = c("references1.bib", "references2.bib")` means that we supply two larger bib files which will be combined and cleaned before the entries matching the citations in `manuscript.Rmd` will be extracted.
- `new_bib_file = "partial_bib.bib"` specifies the name of the newly created (partial) `.bib` files. If we, as in the current examples, request separate `.bib` files for different sections of the document, these will by default be saved by adding `_1`, `_2`, etc. before the file extensions. As we request separate bibliographies for the main text and the appendix, the files will be stored as `partial_bib_1.bib` and `partial_bib_2.bib`. Note that we must already supply matching file names for the bibliographies in our YAML header.
- `by_sections = c("<!--- appendix split -->")` defines the split point included in our `.Rmd` script. Here, we only supply one split point, which means that `tidy_bib()` will extract citations separately before and after the split point (and thus produce two separate `.bib` files). Note that we can easily add split points, e.g., by including distinct comments in the `.Rmd` script and adding them to the input vector for the `by_sections` argument.
- `repair`, `replace_curly_braces`, `removeISSN`, `removeISBN`, `removeDOI`, and `remove URL` are additional options that define how we want to tidy up our new `.bib` file(s).

Document body

```
## Main Text

This is the main text. It cites one paper by @Athey2019-fy inline.
We also cite another one by @Arceneauxetal2012.

Let's also cite a paper in parantheses [@Loges2001-bh].
Here is also some more intricate stuff with prefixes and suffixes
[e.g., @Hargittai2008-fa; @AllcottGentzkow2017, pp. 31-57].

## References
<div id="refs_main"></div>

\newpage

<!--- appendix split -->
## Appendix

This is the appendix. It cites some other paper [@Friedman2009-gx].

## References
<div id="refs_app"></div>
```

In the document body, we can then freely include citations as we usually would. When it comes to printing our bibliographies, there are two things that we handle slightly differently from the default way of including bibliographies in Markdown.

1. `<!--- tidy_bib appendix split -->`, supplied as a Markdown comment, defines the split point.
2. We override the default of printing bibliograpfies at the end of the document by adding `<div id="refs_main"></div>` and `<div id="refs_app"></div>`, respectively. This ensure that the two bibliographies will be printed where we want them to be printed. Note that the names of the arguments must match those defined in the YAML header, just like the file names of the corresponding `.bib` files must match those of the new `.bib` files produced by `tidy_bib`.

On the following page, you can see the knitted PDF document generated from `manuscript.Rmd`.

Main Text

This is the main text. It cites one paper by Athey, Tibshirani, and Wager (2019) inline. We also cite another one by Arceneaux, Johnson, and Murphy (2012).

Let's also cite a paper in parantheses (Loges and Jung 2001). Here is also some more intricate stuff with prefixes and suffixes (e.g., Hargittai and Walejko 2008; Allcott and Gentzkow 2017, 31–57).

References

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31 (2): 211–36.

Arceneaux, Kevin, Martin Johnson, and Chad Murphy. 2012. "Polarized Political Communication, Oppositional Media Hostility, and Selective Exposure." *Journal of Politics* 74 (1): 174–86.

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. "Generalized Random Forests." *Ann. Stat.* 47 (2): 1148–78.

Hargittai, Eszter, and Gina Walejko. 2008. "The Participation Divide: Content Creation and Sharing in." *Inf. Commun. Soc.* 11 (2): 239–56.

Loges, William E, and Joo-Young Jung. 2001. "Exploring the Digital Divide: Internet Connectedness and Age." *Communic. Res.* 28 (4): 536–62.

Appendix

This is the appendix. It cites some other paper (Friedman, Hastie, and Tibshirani 2009).

References

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2009. *The Elements of Statistical Learning*. Vol. 1. New York: Springer series in statistic.