

Citations in Markdown Part 3

Martin Fenner, Gobbledygook

June 24, 2013

After the post last week and the crazy discussion that followed I would understand that you feel you have heard enough about citations in markdown. But I had the feeling last week that something was still missing, and I have done some more thinking. What we have so far:

- Pandoc has nice support for citations, including Citation Style Language support (i.e. it is using the same 5000+ citation styles as Zotero, Mendeley and Papers).
- Pandoc requires a separate file to store the citations, typically in bibtex format. This is fine for some people, but can make the workflow complicated for short documents or when several people work on the bibliography at the same time.
- Citations are similar to links, and we can use links for almost all the functionality we need, making it much easier to add citations to a text. The problem is a) citations that don't include a weblink, b) being able to do this offline, and c) where in the HTML to store the citation metadata.

And I looked at how Wikipedia is doing this, and they use a) links, b) citations and c) footnotes. If Wikipedia thinks that it can't do without citations and do everything as links, then maybe we also shouldn't enforce this for scholarly texts.

I think what we need is the best of both worlds. We should use the Pandoc citation workflow, as it is similar to what we are used to from other authoring environments, and we get good citation style support, including more complex formatting of references. Some reference managers already support copy/paste of Pandoc citation keys. The inclusion of a bibtex file with a scholarly markdown text is also a bonus, as it allows the automated extraction of citations, e.g. by manuscript submission systems.

We also want to support a simpler solution for shorter texts or when people don't want to use a separate bibtex file. Here we would add the citations as links, ideally in a syntax very similar to Pandoc citation keys:

Johnson [Johnson2006] didn't agree with ...

[Johnson2006]: <http://dx.doi.org/10.1002/aris.201> "Data sharing in the sciences"

We need to write a tool that parses the markdown before Pandoc, fetches the citation metadata for these links in bibtex format (e.g. using CrossRef Content Negotiation), and adds them to the existing bibtex file (or creates a new bibtex file). The next time the markdown is parsed, the citation is already “cached” in the bibtex file. Those people who don’t have such a tool would see the citation as link (???), with the essential information (DOI or URL) preserved so that a downstream tool can fetch the bibliographic information. Some people were worried about typos in DOIs and URLs. They can add additional information - e.g. the title of the paper - in double quotes to allow checking of the correct DOI.

This workflow now makes a lot of sense to me, as it uses existing solutions, but also allows for easy entering of citation information in a way similar to the knetcitations and kcite tools. As I use jekyll and am a Ruby developer, I will implement the citation parsing as a jekyll plugin.