A Call for Scholarly Markdown

Martin Fenner, Gobbledygook

December 13, 2012

Markdown is a lightweight markup language, originally created by John Gruber for writing content for the web. Other popular lightweight markup languages are Textile and Mediawiki. Whereas Mediawiki markup is of course popular thanks to the ubiquitous Wikipedia, Markdown seems to have gained momentum among scholars. Markdown really focuses on writing content, many of the features of today's word processors are just a distraction (e.g. fonts, line spacing or style sheets). Adding markup for document structure (e.g. title, authors or abstract) on the other hand is overly complicated with tools such as Microsft Word.

Fortunately or unfortunately there are several versions (or flavors) of Markdown. The original specification by John Gruber hasn't been updated for years. Github uses Markdown with some minor modifications. Multimarkdown and Pandoc provide features important for scholarly content, e.g. citations, superscript and tables.

- Markdown
- Github-flavored Markdown
- Multimarkdown
- Pandoc

The Pandoc flavor of Markdown probably comes closest to the requirements of a scholar, but still has limitations, e.g. support for metadata and tables isn't very flexible. I propose that we as a community create a new Scholarly Markdown flavor, which takes into account most of the use cases important for scholarly content.

One of the big advantages of Markdown is that the format can not only be translated to HTML, but also to other formats, and Pandoc is particularly good in translating to and from many different formats. We want to make sure that Scholarly Markdown not only translates into nice Scholarly HTML (with good support for HTML5 tags relevant for scholars), but also into Microsot Word, LaTeX and PDF, as these are the formats typically required by manuscript tracking systems.

Some of the features required for Scholarly Markdown include:

- Superscript and subscript

- Highlighting text (supporting the HTML tag `<mark>`)
- Captions for tables and figures (with support for the HTML tags `<caption>` and `<figcaption>`)
- Support for document sections (the HTML5 tags `<article>`, `<header>`, `<footer>`, `<section>`)
- Good table support
- Math support
- Good citation support
- Support for comments and annotations

Multimarkdown and Pandoc of course already support many of these features. Tables and citations are two examples where it is important to not only support them, but support them in a non-intrusive way that doesn't get in the way of the flow of writing.

BTW, this wouldn't be the first community flavor for Markdown. The screenwriting community has done this already with Fountain.