Turn your Blog into an eBook

Posted: 05 Mar 2012 02:44:38

If you have worked through the exhausting procedure of converting your blog to LaTeX: see posts (1), (2) and (3), you will be glad to hear that turning your blog into an image free eBook is *almost effortless.* In this post I will describe how I convert my blog into EPUB and MOBI eBooks.

eBooks how the cool kids are reading eBook readers like Kindles, Nooks, iPads and many cell phones are optimized for plain old prose. They excel at displaying reflowable text in a variety of fonts, sizes and styles. One eBook reader feature, dear to my old fart eyes, is the ability to increase the size of text. All eBooks are potentially large print editions. There are other advantages: most readers can store hundreds, if not thousands of books, making them portable libraries. It's now technically possible to hand a kindergarten student a little tablet that holds every single book he will use from preschool to graduate school. The only obstacle is the rapacious textbook industry and their equally rapacious eBook publishing enablers. But fear not open source man will save the day. *The days of overpriced digital goods are over!* I will never pay more than a few bucks for an eBook because I can make my own and so can you! Let's get together and kill off another industry that so has it coming!

PDFs, EPUBs and MOBIs Native eBook file formats like EPUB and MOBI do not handle complex page layouts well. If your document contains a lot of mathematics, figures and well placed illustrations stick with PDF workflows.[1] You will save yourself and your readers a lot of grief. But, if your document is a prose masterpiece, a veritable great American novel, then "publishing" it as an EPUB or MOBI is a great way to target eBook readers. EPUBs and MOBIs can be compiled from many sources. I start with the LaTeX files I created for the PDF version of this blog because I hate doing the same boring task twice. By far the most time-consuming part of converting WordPress export XML to LaTeX is editing the pandoc generated `*.tex` files to resolve figures and fix odd run-together-words and paragraphs. To preserve these edits I use pandoc to convert my edited `*.tex` to `*.markdown` files.

Markdown Markdown is a very simple text oriented format. A markdown file is completely readable exactly the way it is. All you need is a text editor. Even text

[1] LaTeX is usually compiled to PDF making it one of hundreds of PDF workflows.

editors are overkill. You could compose markdown with early 20th century mechanical typewriters; it's a low tech format for the ages: perfect for prose.

The J verb `MarkdownFrLatex`[2] calls pandoc and converts my `*.tex` files to `*.markdown`. I place my markdown in the directory

```
c:/pd/blog/wp2epub
```

and to track changes to my markdown files I GIT this directory. `MarkdownFrLatex` strips out image inclusions and removes typographic flourishes. When it succeeds it writes a simple markdown file and when it fails it writes a `*.baddown` file. Baddown files are `*.tex` files that contain `lstlistings` and complex figure environments that are best resolved with manual edits. After removing such problematic LaTeX environments the J verb `FixBaddown` calls pandoc and turns baddown files into markdown files.

Generating EPUB and MOBI files When the conversion to markdown is complete I run `MainMarkdown` to mash all my files into one large markdown file with an eBook header. The eBook header for this blog is:

```
% Analyze the Data not the Drivel
% John D. Baker
```

The first few lines of the consolidated `bm.markdown` file are:

```
% Analyze the Data not the Drivel
% John D. Baker

#['Whats In it for
Facebook?](http://bakerjd99.wordpress.com/2009/09/05/whats-in-it-for-facebook/)

------------------------------------------------------------------------

*Posted: 05 Sep 2009 22:44:50*

[Facebook](http://www.facebook.com) is huge: they brag about a user
count well north of one hundred million. If only 0.5% of their users are
active 'thats 500,000 *concurrent users.* How many expensive servers
does it take to support such a load? .....
```

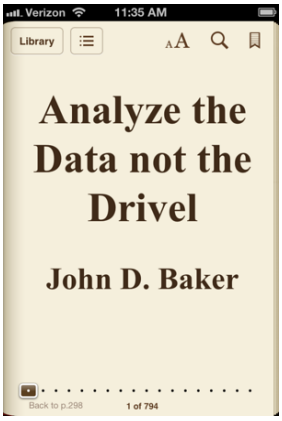
Generating an EPUB from `bm.markdown` is a simple matter of opening up your favorite

[2] All the J verbs referenced in this post are in the script `TeXfrWpxml.ijs`

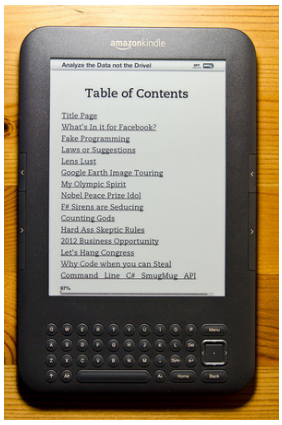
command line shell and issuing the pandoc command:

```
pandoc -S --epub-cover-image=bmcover.jpg -o bm.epub bm.markdown
```

You can read the resulting EPUB file `bm.epub` on any EPUB eBook reader. Here's a screen shot of `bm.epub` on my iPhone.



iPhone loaded with `bm.epub`



Kindle loaded with `bm.mobi`

The last step converts `bm.epub` to `bm.mobi`. MOBI is a native Kindle format. Pandoc can generate MOBI from `bm.markdown` but it inexplicably omits a table of contents. *No problemo:* I use Calibre to convert `bm.epub` to `bm.mobi`. Calibre properly converts the embedded EPUB table of contents to MOBI. Here's `bm.mobi` on a Kindle.

All the "published" versions of this blog are available on the ***Download this Blog*** page so please help yourself!

> *From the blog: Analyze the Data not the Drivel*
> *John D. Baker — revised: August 14, 2020*