What Does Google Want Us to Do With All These Free PDF eBooks?

Richard Crocker Editor

August 31, 2006

Editor's Note: ARTS PDF and Planet PDF are both divisions of BinaryThing

If you haven't heard yet, Google has just announced that a bunch of the books it has been scanning for its Google Books Library Program are now available for free download as PDFs. There's no doubt Google needs to be applauded for the idea, but the execution (i.e. the books they've produced) could definitely do with some work. The PDF books are **difficult to download**, **large** in size, of such **low resolution** they're difficult to read, **unsearchable**, and **do not allow the user to copy text** from them. It's left me wondering what Google expects people to do with the books.

Some of the issues with the PDF books are minor and can be fixed in no time. Others are essentially determined by Google's policy and real plan behind these free books. The key question Google needs to ask itself: does it want these books to be really useful and usable, or does it just want to give away a hell of a lot of pages of free content that readers do little with?

Planet PDF has been giving free eBooks (like Tolstoy, Aesop, Dickens, etc.) away for years, and a key part of the philosophy behind doing it was to share free books that actually enhanced the reading experience off and on a computer (in particular, via the rich array of functionality PDF files can support). Of course, our eBooks are far from perfect (I'm not going to detail here why), but they do make use of a bunch of PDF settings, additions, and optimizations that Google could learn from. Below I'll highlight a few issues I think Google should look at to make better books.

Google Should Think About

Allowing users to save directly to their computer

Clicking on a web link to a PDF file normally by default opens the document inside a Web browser -- PDF viewers inside Web browsers are notoriously unreliable in my experience (proven again as I researched this article with my browser and/or PDF viewer crashing) so you should always allow users to save the file directly to disk (in Windows you right-click and choose to Save As). As it is, Google's download button means PDFs

will just load up in your browser, from which you can then save. It took me a few attempts to get my first free book as it wouldn't load in my Web browser.

Optimizing PDFs for use on the Web (a.k.a. 'Fast Web View' or Linearization)

This is something Google should fix immediately as all users will be experiencing problems with it. The good thing is it's trivial to run big collections of PDFs through the optimization process -- both Acrobat and our Nitro PDF Professional do it. Optimization is a technique that allows byte-serving, which means the pages in PDF documents can be downloaded one at a time. So , for example, if you load a 300 page PDF in your web browser, you can start reading page one well before page 100 has been downloaded.

Fixing Low Resolution/Readability

My eyes aren't great -- they're not too bad -- but I reckon a person with 20/20 vision would still find these books a strain on the eyes. Google's reason for such low resolution may well be to keep the file sizes down but they seem to have gone too far. I'd suggest offering both a high and low resolution version so readers who have fast internet, or those that really need a quality version of the book can get what they need. Again, this isn't hard to implement, it's just another automated step in the process of producing the final PDF book.

Can I search, Google?

Would you be surprised to learn that the books produced by Google cannot be searched? Yep, true. This is where I really don't understand the philosophy behind them giving these books away free. Clearly they have been OCR'd (optical character recognition'd), that is, the scanned pages have had the text converted to indexable, searchable text (after all, you can search them on the Web), yet they can't be searched when you download them.

There are many software programs out there that can create PDFs that retain the original look of the pages and then have text beneath the page that can be searched, selected and copied. It's simply a matter of Google making this an extra step in the PDF creation process. Virtually all eBooks online are searchable, including all 19,000 free books at Project Gutenberg.

Copying and Reusing Text/Highlighting and Annotating Text

If students are one of the intended audiences then selecting and copying text seems important to me -- there appears to be no reason why readers shouldn't be allowed to grab a paragraph or two of text from a public domain work instead of having to rewrite it when they want to quote the text in an essay. This again comes back to the fact that Google has chosen not to offer PDF books that have been OCR'd, meaning the text is not searchable or selectable.

Perhaps less important for the average punter (the everyday person) who isn't in to marking up and commenting on the text in a PDF file, but, for many of us using highlighting tools to keep track of important content in documents and at times adding our own comments to text via pop-up notes, being able to select and copy text is essential to our research and review process.

Bookmarks

Most Planet PDF regulars will know about bookmarks and how handy they can be -- we use them in our free eBooks and they're a great navigation aid for getting to the start of any chapter fast and work much like a table of contents.

It's trivial to set up PDFs so they automatically display the list of bookmarks next to the pages. The issue is that to automate the process of creating bookmarks all text-based content in a PDF needs to be actual text rather than pictures or images of text (as the Google books currently are). If Google converted all pages to text (using OCR technology) then it would be possible to create bookmarks from this automatically. There are plenty of tools out there to create bookmarks based on font sizes and styles.

In Summary

Google, do you really want us to read and use these books?