02-15-2023, 19:43 #8

BetterRed
null operator

Posts: 18,972
Karma: 22381319
Join Date: Mar 2012
Location: Sydney Australia
Device: none

Quote:

> Originally Posted by **sgmoore**
> *It's definitely faster (a quick one off test shows that spawning ebook-convert is about five times slower).*
>
> *Unfortunately it is not better and indeed not good enough. I have some files which look like they have been **generated as epub files by Microsoft Word,** and the count_pages algorithm produces text which is about four times larger than ebook-convert. (A quick glance shows thousands of font-family entries which have not been removed by count_pages).*

AFAIK Microsoft Word itself is incapable of creating EPUB files directly.

The ePUBTools Word addin can create EPUBs from within Word, and there are a number of tools, including calibre, that will convert MS Word's native format DOCX files to EPUB.

Added: InDesign is a more likely candidate as the source of poorly formed EPUBs.

BR

Last edited by BetterRed; 02-15-2023 at 19:57.

KARMA ! QUOTE

02-16-2023, 00:41 #9

kovidgoyal
creator of calibre

Posts: 42,259
Karma: 19574690
Join Date: Oct 2006
Location: Mumbai, India
Device: Various

If you care about speed use the extract_text() function from calibre.db.fts.text

Support my calibre work

KARMA ! QUOTE

02-16-2023, 13:00 #10

sgmoore
Member

Posts: 18
Karma: 10
Join Date: Mar 2021
Device: Kindle Voyage

Quote:

> Originally Posted by **kovidgoyal**
> *If you care about speed use the extract_text() function from calibre.db.fts.text*

Thank you. That looks to do what I want. It does not produce the exact same results as ebook-convert but ignoring formatting and white-space issues its is extremely close. I tried it on about 1000 books and the worse case was still 99% similar and the vast majority of them were over 99.9% similar.

Also only takes about 1/20 of the time to call ebook-convert.

Thanks again.

KARMA ! QUOTE

02-16-2023, 21:22 #11

Quoth
the rook, bossing Never.

Calibre converting docx to epub works better than the plugins I've tried on Word & LO Writer. It seems to work better than Indesign and other commercial tools judging by the commercial ebooks from big publishers.
Indesign should only be used for fancy colour coffee table books and glossy magasines.

Calibre is jjust about perfect from properly formatted docx (made by Word or extra Save As in LO