Demonstration of DOCX support in calibre

This document demonstrates the ability of the calibre DOCX Input plugin to convert the various typographic features in a Microsoft Word (2007 and newer) document. Convert this document to a modern ebook format, such as AZW3 for Kindles or EPUB for other ebook readers, to see it in action.

There is support for images, tables, lists, footnotes, endnotes, links, dropcaps and various types of text and paragraph level formatting.

To see the DOCX conversion in action, simply add this file to calibre using the **"Add Books"** button and then click "**Convert".** Set the output format in the top right corner of the conversion dialog to EPUB or AZW3 and click **"OK"**.

Text Formatting

Inline formatting

Here, we demonstrate various types of inline text formatting and the use of embedded fonts.

Here is some **bold,** *italic, **bold-italic,*** <u>underlined</u> and ~~struck out~~ text. Then, we have a super^{script} and a sub_{script}. Now we see some red, green and blue text. Some text with a <mark>yellow highlight</mark>. Some text in a box. Some text in inverse video.

A paragraph with styled text: *subtle emphasis* followed by **strong text** and ***intense emphasis***. This paragraph uses document wide styles for styling rather than inline text properties as demonstrated in the previous paragraph — calibre can handle both with equal ease.

Fun with fonts

This document has embedded the Ubuntu font family. The body text is in the Ubuntu typeface, here is `some text in the Ubuntu Mono typeface, notice how every letter has the same width, even i and m`. Every embedded font will automatically be embedded in the output ebook during conversion.

Paragraph level formatting

You can do crazy things with paragraphs, if the urge strikes you. For instance this paragraph is right aligned and has a right border. It has also been given a light gray background.

For the lovers of poetry amongst you, paragraphs with hanging indents, like this often come in handy. You can use hanging indents to ensure that a line of poetry retains its individual identity as a line even when the screen is too narrow to display it as a single line. Not only does this paragraph have a hanging indent, it is also has an extra top margin, setting it apart from the preceding paragraph.

Tables

ITEM	NEEDED
Books	1
Pens	3
Pencils	2
Highlight er	2 colors
Scissors	1 pair

Tables in Word can vary from the extremely simple to the extremely complex. calibre tries to do its best when converting tables. While you may run into trouble with the occasional table, the vast majority of common cases should be converted very well, as demonstrated in this section. Note that for optimum results, when creating tables in Word, you should set their widths using percentages, rather than absolute units. To the left of this paragraph is a floating two column table with a nice green border and header row.

Now let's look at a fancier table—one with alternating row colors and partial borders. This table is stretched out to take 100% of the available width.

City or Town	Point A	Point B	Point C	Point D	Point E
Point A	—				
Point B	87	—			
Point C	64	56	—		
Point D	37	32	91	—	
Point E	93	35	54	43	—

Next, we see a table with special formatting in various locations. Notice how the formatting for the header row and sub header rows is preserved.

College	New students	Graduating students	Change
Undergraduate			
Cedar University	110	103	+7
Oak Institute	202	210	-8
Graduate			
Cedar University	24	20	+4
Elm College	43	53	-10
Total	998	908	90

Source: Fictitious data, for illustration purposes only

Next, we have something a little more complex, a nested table, i.e. a table inside another table. Additionally, the inner table has some of its cells merged. The table is displayed horizontally centered.

One	Two		To the left is a table inside a table, with some cells merged.
Three	Four		

We end with a fancy calendar, note how much of the original formatting is preserved. Note that this table will only display correctly on relatively wide screens. In general, very wide tables or tables whose cells have fixed width requirements don't fare well in ebooks.

December 2007

Sun	Mon	Tue	Wed	Thu	Fri	Sat
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

Structural Elements

Miscellaneous structural elements you can add to your document, like footnotes, endnotes, dropcaps and the like.

Footnotes & Endnotes

Footnotes[1] and endnotes are automatically recognized and both are converted to endnotes, with backlinks for maximum ease of use in ebook devices.

Dropcaps

Drop caps are used to emphasize the leading paragraph at the start of a section. In Word it is possible to specify how many lines of text a drop-cap should use. Because of limitations in ebook technology, this is not possible when converting. Instead, the converted drop cap will use font size and line height to simulate the effect as well as possible. While not as good as the original, the result is usually tolerable. This paragraph has a "D" dropcap set to occupy three lines of text with a font size of 58.5 pts. Depending on the screen width and capabilities of the device you view the book on, this dropcap can look anything from perfect to ugly.

Links

Two kinds of links are possible, those that refer to an external website and those that refer to locations inside the document itself. Both are supported by calibre. For example, here is a link pointing to the calibre download page. Then we have a link that points back to the section on paragraph level formatting in this document.

Table of Contents

There are two approaches that calibre takes when generating a Table of Contents. The first is if the Word document has a Table of Contents itself. Provided that the Table of Contents uses hyperlinks, calibre will automatically use it. The levels of the Table of Contents are identified by their left indent, so if you want the ebook to have a multi-level Table of Contents, make sure you create a properly indented Table of Contents in Word.

If no Table of Contents is found in the document, then a table of contents is automatically generated from the headings in the document. A heading is identified as something that has the Heading 1 or Heading 2, etc. style applied to it. These headings are turned into a Table of Contents with Heading 1 being the topmost level, Heading 2 the second level and so on.

---

[1] In paged media, footnotes are usually displayed at the bottom of the text. However, in ebooks, a better paradigm is to make them clickable endnotes that the user can browse at her pleasure. This conversion is handled automatically by calibre.

You can see the Table of Contents created by calibre by clicking the Table of Contents button in whatever viewer you are using to view the converted ebook.

Images

Images can be of three main types. Inline images are images that are part of the normal text flow, like this image of a green dot ●. Inline images do not cause breaks in the text and are usually small in size. The next category of image is a floating image, one that "floats " on the page and is surrounded by text. Word supports more types of floating images than are possible with current ebook technology, so the conversion maps floating images to simple left and right floats, as you can see with the left and right arrow images on the sides of this paragraph.

The final type of image is a "block" image, one that becomes a paragraph on its own and has no text on either side. Below is a centered green dot.

Centered images like this are useful for large pictures that should be a focus of attention.

Generally, it is not possible to translate the exact positioning of images from a Word document to an ebook. That is because in Word, image positioning is specified in absolute units from the page boundaries. There is no analogous technology in ebooks, so the conversion will usually end up placing the image either centered or floating close to the point in the text where it was *inserted*, not necessarily where it appears on the page in Word.

Lists

All types of lists are supported by the conversion, with the exception of lists that use fancy bullets, these get converted to regular bullets.

Bulleted List

- One
- Two

Numbered List

1. One, with a very long line to demonstrate that the hanging indent for the list is working correctly
2. Two

Multi-level Lists

1. One
   1.1. Two
      1.1.1. Three
      1.1.2. Four with a very long line to demonstrate that the hanging indent for the list is working correctly.
      1.1.3. Five
2. Six

A Multi-level list with bullets:

- One
  - Two
    - This bullet uses an image as the bullet item
      o Four
- Five

Continued Lists

i. One
ii. Two

An interruption in our regularly scheduled listing, for this essential and very relevant public service announcement.

iii. We now resume our normal programming
iv. Four