Microsoft Windows [Version 6.2.9200](c) 2012 Microsoft Corporation. All rights reserved.  
D:\git\_pdf\pdf2htmlEX>pdf2htmlEX

**Usage: pdf2htmlEX [options] <input.pdf> [<output.html>]**

-f,--first-page <int> first page to convert (default: 1)  
 -l,--last-page <int> last page to convert (default: 2147483647)  
 --zoom <fp> zoom ratio  
 --fit-width <fp> fit width to <fp> pixels  
 --fit-height <fp> fit height to <fp> pixels  
 --use-cropbox <int> use CropBox instead of MediaBox (default: 1)  
 --hdpi <fp> horizontal resolution for graphics in DPI (default: 144)  
 --vdpi <fp> vertical resolution for graphics in DPI (default: 144)  
 --embed <string> specify which elements should be embedded into output  
 --embed-css <int> embed CSS files into output (default: 1)  
 --embed-font <int> embed font files into output (default: 1)  
 --embed-image <int> embed image files into output (default: 1)  
 --embed-javascript <int> embed JavaScript files into output (default: 1)  
 --embed-outline <int> embed outlines into output (default: 1)  
 --split-pages <int> split pages into separate files (default: 0)  
 --dest-dir <string> specify destination directory (default: ".")  
 --css-filename <string> filename of the generated css file (default: "")  
  
 --page-filename <string> filename template for splitted pages (default:"")  
 --outline-filename <string> filename of the generated outline file (default: "")  
 --process-nontext <int> render graphics in addition to text (default: 1)  
  
 --process-outline <int> show outline in HTML (default: 1)  
 --printing <int> enable printing support (default: 1)  
 --fallback <int> output in fallback mode (default: 0)  
 --embed-external-font <int> embed local match for external fonts (default: 1)  
 --font-suffix <string> suffix for embedded font files (.ttf,.otf,.woff,.svg) (default: ".ttf")  
 --decompose-ligature <int> decompose ligatures, such as 詮?-> fi (default:0)  
 --auto-hint <int> use fontforge autohint on fonts without hints (default: 0)  
 --external-hint-tool <string> external tool for hinting fonts (overrides --auto-hint) (default: "")  
 --stretch-narrow-glyph <int> stretch narrow glyphs instead of padding them (default: 0)  
 --squeeze-wide-glyph <int> shrink wide glyphs instead of truncating them (default: 1)  
 --override-fstype <int> clear the fstype bits in TTF/OTF fonts (default: 0)  
 --heps <fp> horizontal threshold for merging text, in pixels (default: 1)  
 --veps <fp> vertical threshold for merging text, in pixels (default: 1)  
 --space-threshold <fp> word break threshold (threshold \* em) (default:0.125)  
 --font-size-multiplier <fp> a value greater than 1 increases the rendering accuracy (default: 4)  
 --space-as-offset <int> treat space characters as offsets (default: 0)  
 --tounicode <int> how to handle ToUnicode CMaps (0=auto, 1=force,-1=ignore) (default: 0)  
 --optimize-text <int> try to reduce the number of HTML elements used for text (default: 0)  
 -o,--owner-password <string> owner password (for encrypted files)  
 -u,--user-password <string> user password (for encrypted files)  
 --no-drm <int> override document DRM settings (default: 0)  
 --clean-tmp <int> remove temporary files after conversion (default: 1)  
 --data-dir <string> specify data directory (default: ".\data")  
 --debug <int> print debugging information (default: 0)  
 -v,--version print copyright and version info  
 -h,--help print usage information

**Command line**

**NAME** pdf2htmlEX - converts PDF to HTML without losing text and format.  
**USAGE** **pdf2htmlEX [options] <input-filename> [<output-filename>]**

**DESCRIPTION**

pdf2htmlEX is a utility that converts PDF files to HTML files.  
pdf2htmlEX tries its best to render the PDF precisely, maintain proper styling, while retaining text and optimizing for Web.  
Fonts are extracted form PDF and then embedded into HTML (Type 3 fonts are not supported). Text in the converted HTML file is usually selectable and copyable.  
Other objects are rendered as images and also embedded.

**OPTIONS**

* **Pages**
  + -f, --first-page <num> (Default: 1)

Specify the first page to process

* + -l, --last-page <num> (Default: last page)

Specify the last page to process

* **Dimensions**
  + --zoom <ratio>, --fit-width <width>, --fit-height <height>

--zoom specifies the zoom factor directly; --fit-width/height specifies the maximum width/height of a page, the values are in pixels. If multiple values are specified, the minimum one will be used.  
If none is specified, pages will be rendered as 72DPI.

* + --use-cropbox <0|1> (Default: 1)

Use CropBox instead of MediaBox for output.

* + --hdpi <dpi>, --vdpi <dpi> (Default: 144)

Specify the horizontal and vertical DPI for images

* **Output**
  + --embed <string>
  + --embed-css <0|1> (Default: 1)
  + --embed-font <0|1> (Default: 1)
  + --embed-image <0|1> (Default: 1)
  + --embed-javascript <0|1> (Default: 1)
  + --embed-outline <0|1> (Default: 1)

Specify which elements should be embedded into the output HTML file. If switched off, separated files will be generated along with the HTML file for the corresponding elements. --embed accepts a string as argument. Each letter of the string must be one of `cCfFiIjJoO`, which corresponds to one of the --embed-\*\*\* switches. Lower case letters for 0 and upper case letters for 1. For example, `--embed cFIJo` means to embed everything but CSS files and out‐lines.

* + --split-pages <0|1> (Default: 0)

If turned on, the content of each page is stored in a separated file. --page-filename may be used to specify the format for the filenames for individual pages. A %d placeholder may be included to indicate where the page number should be placed. The placeholder supports a limited subset of normal numerical placeholders, including specified width and zero padding.

If --page-filename does not contain a placeholder for the page number, the page number will be inserted directly before the file extension. If the filename does not have an extension, the page number will be placed at the end of the file name.

If --page-filename is not specified, <input-filename> will be used for the output filename, replacing the extension with .page and adding the page number directly before the extension.

This switch is useful if you want pages to be loaded separately & dynamically -- a supporting server might be necessary.

Examples

* + - * pdf2htmlEX --split-pages 1 foo.pdf  
        Yields page files foo1.page, foo2.page, etc.
      * pdf2htmlEX --split-pages 1 foo.pdf bar.baz  
        Yields page files bar1.baz, bar2.baz, etc.
      * pdf2htmlEX --split-pages 1 foo.pdf page%dbar.baz  
        Yields page files page1bar.baz, page2bar.baz, etc.
      * pdf2htmlEX --split-pages 1 foo.pdf bar%03d.baz  
        Yields page files bar001.baz, bar002.baz, etc.
  + --dest-dir <dir> (Default: .)

Specify destination folder.

* --css-filename <filename> (Default: <none>)

Specify the filename of the generated css file, if not embedded.  
If it's empty, the file name will be determined automatically.

* --page-filename <filename> (Default: <none>)

Specify the filename template for pages. This is only useful when --split-pages is 1  
If it's empty, a default one will be used, see description of --split-pages

* --outline-filename <filename> (Default: <none>)

Specify the filename of the generated outline file, if not embedded.  
If it's empty, the file name will be determined automatically.

* --process-nontext <0|1> (Default: 1)

Whether to process non-text objects (as images)

* --process-outline <0|1> (Default: 1)

Whether to show outline in the generated HTML

* --printing <0|1> (Default: 1)

Enable printing support. Disabling this option may reduce the size of CSS.

* --fallback <0|1> (Deafult: 0)

Output in fallback mode, for better accuracy and browser compatibility, but the size becomes larger.

* **Fonts**
  + --embed-external-font <0|1> (Default: 1)

Specify whether the local matched fonts, for fonts not embedded in PDF, should be embedded into HTML.  
If this switch is off, only font names are exported such that web browsers may try to find proper fonts themselves, and that might cause issues about incorrect font metrics.

* --font-suffix <suffix> (Default: .ttf)

Specify the suffix of fonts extracted from the PDF file.

* --decompose-ligature <0|1> (Default: 0)

Decompose ligatures. For example 'fi' -> 'f''i'.

* --auto-hint <0|1> (Default: 0)

If set to 1, hints will be generated for the fonts using fontforge.  
This may be preceded by --external-hint-tool.

* --external-hint-tool <tool> (Default: <none>)

If specified, the tool will be called in order to enhanced hinting for fonts, this will precede --auto-hint.  
The tool will be called as '<tool> <in.suffix> <out.suffix>', where suffix will be the same as specified for --font-suffix.

* --stretch-narrow-glyph <0|1> (Default: 0)

If set to 1, glyphs narrower than described in PDF will be stretched; otherwise space will be padded to the right of the glyphs

* --squeeze-wide-glyph <0|1> (Default: 1)

If set to 1, glyphs wider than described in PDF will be squeezed; otherwise it will be truncated.

* **Text**
  + --heps <len>, --veps <len> (Default: 1)

Specify the maximum tolerable horizontal/vertical offset (in pixels).  
pdf2htmlEX would try to optimize the generated HTML file moving Text within this distance.

* + --space-threshold <ratio> (Default: 0.125)

pdf2htmlEX would insert a whitespace character ' ' if the distance between two consecutive letters in the same line is wider than ratio \* font\_size.

* + --font-size-multiplier <ratio> (Default: 4.0)

Many web browsers limit the minimum font size, and many would round the given font size, which results in incorrect rendering. Specify a ratio greater than 1 would resolve this issue, however it might freeze some browsers.  
For some versions of Firefox, however, there will be a problem when the font size is too large, in which case a smaller value should be speci‐fied here.

* + --space-as-offset <0|1> (Default: 0)

If set to 1, space characters will be treated as offsets, which allows a better optimization.  
For PDF files with bad encodings, turning on this option may cause losing characters.

* + --tounicode <-1|0|1> (Default: 0)

A ToUnicode map may be provided for each font in PDF which indicates the 'meaning' of the characters. However often there is better "ToUnicode" info in Type 0/1 fonts, and sometimes the ToUnicode map provided is wrong. If this value is set to 1, the ToUnicode Map is always applied, if provided in PDF, and characters may not render correctly in HTML if there are collisions.  
If set to -1, a customized map is used such that rendering will be correct in HTML (visually the same), but you may not get correct characters by select & copy & paste.  
If set to 0, pdf2htmlEX would try its best to balance the two methods above.

* + --optimize-text <0|1> (Deafult: 0)

If set to 1, pdf2htmlEX will try to reduce the number of HTML elements used for text. Turn it off if anything goes wrong.

* **PDF Protection**
  + -o, --owner-password <password>

Specify owner password

* + -u, --user-password <password>

Specify user password

* + --no-drm <0|1> (Default: 0)

Override document DRM settings

* **Misc.**
  + --clean-tmp <0|1> (Default: 1)

If switched off, intermediate files won't be cleaned in the end.

* + --data-dir <dir> (Default: /usr/share/pdf2htmlEX)

Specify the folder holding the manifest and other files (see below for the manifest file)`

* + --css-draw <0|1> (Default: 0)

Experimental and unsupported CSS drawing

* + --debug <0|1> (Default: 0)

Print debug information.

* **Meta**
  + -v, --version

Print copyright and version info

* + --help

Print usage information

**MANIFEST and DATA-DIR**

When split-pages is 0, the manifest file describes how the final html page should be generated.  
By default, pdf2htmlEX will use the manifest in the default data-dir (run `pdf2htmlEX -v` to check), which gives a simple demo of its syntax.  
You can modify the default one, or you can create a new one and specify the correct data-dir in the command line.  
When single-html is 1, all files referred by the manifest must be located in the data-dir.

**EXAMPLE**

pdf2htmlEX /path/to/file.pdf

Convert file.pdf into file.html

pdf2htmlEX --clean-tmp 0 --debug 1 /path/to/file.pdf

Convert file.pdf and leave all intermediate files.

pdf2htmlEX --dest-dir out --single-html 0 /path/to/file.pdf

Convert file.pdf into out/file.html and leave font/image files separated.