User's Guide » Fonts & Languages

Arabic (RTL) text

If you are writing a document in Arabic or Hebrew text (RTL or right-to-left script) and specify *ar* or *he* as the language code, then the layout of block elements is reversed and default alignments are set to right rather than left.
The document is set to RTL *state.* (Internally this is represented by the variable *$directionality*='rtl')

Document State - RTL versus LTR
When the *document is set to* RTL *state*, tables, lists, columns, text alignment and table cell alignment are changed to right-alignment.

mPDF can be set to this state - automatically by the defined language (*ar* or *he*):

```
$mpdf = new mPDF('ar');
```

or by setting it directly:

```
$mpdf = new mPDF('');
$mpdf->SetDirectionality('rtl');
```

Text Bidirectionality
mPDF checks every chunk of text (defined as sections of text between HTML tags) for the presence of any characters in the Unicode ranges \x0590-\x07BF and \xFB50-\xFDFF. If RTL text is detected:

- If the document is RTL, and any of the chunk contains any RTL characters, the letters in each word, **and** the word order in the chunk are both reversed
- If the document is LTR and the chunk contains exclusively RTL characters, then the letters **and** the word order is reversed
- If the document is LTR and the chunk contains mixed RTL & LTR characters, only the letter order is reversed.

> **Note:** *Text Bidirectionality* was changed in mPDF 4.0 so that all text is **always** checked for RTL characters.

Fonts

Arabic is a complex script requiring processing before output. However any appropriate font can be used - as long as it contains the characters in Unicode blocks 'Arabic Presentation Forms' A and B (U+FB50 - U+FDFF, U+FE70 - U+FEFE). Note that quite a large number of fonts contain the isolated characters but not the presentation forms.

2 fonts are bundled with mPDF: XB Zar and XB Riyaz. These are 2 of a number of fonts available from http://wiki.irmug.com/index.php/X_Series_2.

> **Note:** The script handling Arabic text (RTL) was rewritten in mPDF 5.5 with improved support for Pashto/Sindhi/Urdu/Kurdish, especially for joining characters and added new presentation forms.

Non-unicode characters
Some characters in Pashto/Sindhi/Urdu/Kurdish do not have Unicode values for the final/initial/medial forms of the characters. However, some fonts include glyphs for these characters "un-mapped" to Unicode (including XB Zar and XB Riyaz, which are bundled with mPDF).

By editing `config_fonts.php` and adding to appropriate fonts:

```
'unAGlyphs' => true,
```

this will force mPDF to use unmapped glyphs. It requires the font file to include a Format 2.0 POST table which references the glyphs by name as e.g. uni067C.med or uni067C.medi

XB Riyaz, XB Zar, Arabic Typesetting (MS), Arial (MS) all contain this table. NB If you want to know if a font file is suitable, you can open a .ttf file in a text editor and search for "uni067C.med" - if it exists, it may work!

Using "unAGlyphs" forces subsetting of fonts, and will not work with SIP/SMP fonts (using characters beyond the Unicode BMP Plane).

mPDF maps these characters to part of the Private Use Area allocated by Unicode U+F500-F7FF. This could interfere with correct use if the font already utilises these codes (unlikely).

Alef Maksura

Detailed note on the Alef Maksura for advanced users:
U+0649 Alef Maksura only normally appears at the end of a word (in Arabic)
Initial and Medial forms exist in Unicode as FBE8 and FBE9 but are not in most fonts
So the final form is set in mPDF to show as FEF0; Initial and medial forms are shown as isolated/final, so that it does at least display.
It seems that Initial and Medial forms are used in Koranic text.
I have left options encoded in *function InitArabic()* if you want to alter - to make it double-joining, it also needs to be added to *$arabPrevLink* as "\xd9\x89"

> **Note:** mPDF deletes Unicode characters: U+200C,U+200D,U+200E,U+200F zero-width joiner/non-joiner, LTR and RTL marks so they will not appear - even though some fonts contain glyphs for these characters.

See Also

- useLang - Specify whether to recognise and support the HTML attribute lang
- SetAutoFont() - Use AutoFont to auto-detect text language in HTML input
- autoFontGroupSize - Specify the text chunk size to group when autodetecting text language
- disableMultilingualJustify - Specify whether to disable text justification in multilingual documents
- lang - Information on mPDF support for the HTML attribute lang

Printed on Wed 05 Aug 2015 12:37:53 GMT +0100 (DST)