Automatic font selection

> **Note:** This describes automatice font selection in mPDF >= v6.

mPDF has two functions which can be used together or separately:

`autoScriptToLang` - marks up HTML text using the lang attribute, based on the Unicode script block in question, and configurable values in `config_script2lang.php`.

`autoLangToFont` - selects the font to use, based on the HTML lang attribute, using configurable values in `config_lang2font.php`.

For automatic font selection, ideally we would choose the font based on the language in use. However it is actually impossible to determine the language used from a string of HTML text. The Unicode script block can be ascertained, and sometimes this tells us the language e.g. Telugu. However, Cyrillic script is used for example in many different languages. So the best we can do is base it on the script used. However, mPDF does this in two stages via the `lang` attribute, because this allows the options of using either of the stages alone or together:

```
<p>English ру́сский язы́к پښتو</p>
```

↓ **autoScriptToLang** (config_script2lang.php) ↓

```
<p>English <span lang="und-Cyrl">ру́сский язы́к</span> <span lang="ps">پښتو</span></p>
```

↓ **autoLangToFont** (config_lang2fonts.php) ↓

```
Uses "lang" to select font, and to determine OTL features applied
```

autoScriptToLang

```
$mpdf->autoScriptToLang = true;
$mpdf->baseScript = 1;
$mpdf->autoVietnamese = true;
$mpdf->autoArabic = true;
```

`$mpdf->baseScript = 1;` tells mPDF which Script to ignore. It is set by default to "1" which is for Latin script. In this mode, all scripts *except* Latin script are marked up with `lang` attribute. To select other scripts as the base, see the file `/classes/ucdn.php`

Using `autoScriptToLang`, mPDF detects text runs based on Unicode script block; using the values in `config_script2lang.php` it then encloses the text run within a span tag with the appropriate language attribute. For many scripts, the language cannot be determined: see the example above which recognises Cyrillic script and marks it up using `und-Cyrl`, which is a valid IETF tag, coding for language="undetermined", script="Cyrillic".

Two optional refinements are added: Vietnamese text can often be recognised by the presence of certain characters which do not appear in other Latin script langauges, and similarly analysis of the text can attempt to distinguish Arabic, Farsi, Pashto, Urdu and Sindhi. If active, the text will then be marked with a specific language tag e.g. "vi", "pa", "ur", "fa" etc.

These features can be disabled or enabled (default) using the variables `$mpdf->autoVietnamese` `$mpdf->autoArabic`, either in `config.php` or at runtime.

autoLangToFont

```
$mpdf->autoLangToFont = true;
```

You can edit the values in `config_lang2font.php` to specify which fonts are used for which `lang`.

Using text with multiple languages

Recommended ways to use multiple languages in mPDF:

1. If you have full control over the HTML, mark-up the text with the `lang` atribute and use CSS (`:lang` selector preferably); this method means that the language information can also be used by OTL for language dependent substitutions.
2. If you have no control over (user) HTML input and want to output faithfully, use both `autoScriptToLang` and `autoLangToFont`

It is preferable not to use `autoScriptToLang` and `autoLangToFont` unless they are necessary: they will result in increased processing time, and OTL tables will not be able to use language dependent substitutions when undefined languages are set e.g "`und-Cyrl`".

> **Note:** As from mPDF 6.0 automatic font selection can be used within an SVG image. Control is separate from that of the rest of the document. For details, see the defined constants at the top of classes/svg.php file.

Printed on Wed 05 Aug 2015 12:04:54 GMT +0100 (DST)