





## COMBINING DIACRITICS

In Unicode, letters with diacritics (e.g. ÁáĂăÄä) are usually represented as a single character e.g. Unicode U+0196 is an A Umlaut. There are 4 blocks in Unicode of diacritics or 'marks' which can be used to combine with adjacent letters: Combining Diacritical Marks (U+0300 - U+036F), Combining Diacritical Marks Supplement (U+1DC0 - U+1DFF), Combining Marks for Symbols(U+20D0 - U+20FF) and Combining Half Marks (U+FE20 - U+FE2F).

Software applications use special positioning information stored in OpenType font files to reposition the diacritic/mark depending on the context. mPDF does not support this repositioning and is dependent on the font design and original placement of the diacritic:

Á á Ä ä Ā ā İ (Precomposed characters: DejaVu Sans Condensed)

Á á Ä ä Ā ā İ (Using diacritics: DejaVu Sans Condensed)

Á á Ä ä Ā ā İ (Arial Unicode MS)

Á á Ä ä Ā ā İ (Times New Roman)

Á´ á´ Ä` ä` Ä¨ ä¨ İ¨ (Courier New)

**It is recommended to use precomposed characters whenever possible with mPDF.**

# Unicode Supplementary Planes

The original Unicode allocated characters between x0000 and xFFFF (65,536 characters). This 'Basic Multilingual Plane' supported most characters in common use, including a large number of Unified Chinese-Japanese-Korean characters (CJK). Later the Unicode standard was extended to 16 Planes.

The first plane (plane 0), the Basic Multilingual Plane (BMP), is where most characters have been assigned so far.

Plane 1, the Supplementary Multilingual Plane (SMP), is mostly used for historic scripts such as Linear B, but is also used for musical and mathematical symbols.

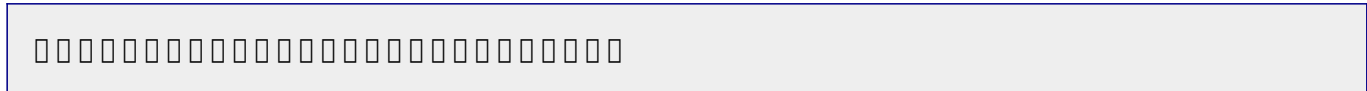
Plane 2, the Supplementary Ideographic Plane (SIP), is used for about 40,000 Unified Han (CJK) Ideographs.

mPDF version 5 supports fonts containing characters from all Unicode Planes. By choosing the correct font, almost every single character from Unicode 5 can be displayed in a PDF file.

## UNICODE SUPPLEMENTARY MULTILINGUAL PLANE (SMP OR PLANE 1) U+10000 - U+1FFFF

### Gothic text

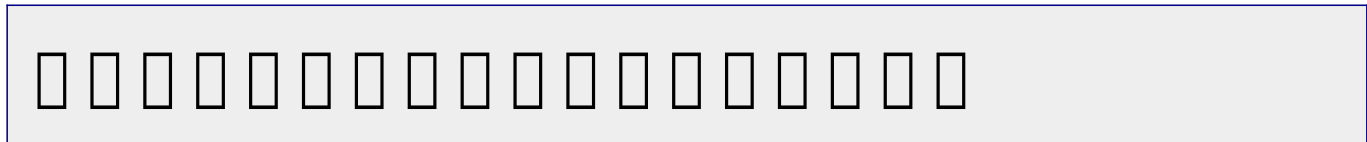
This paragraph shows Gothic text. These characters lie in the Unicode Supplementary Multilingual Plane U+10330 - U+1034F.



Font: MPH2BDamase (damase\_v.2.ttf) available from:  
[http://www.wazu.jp/gallery/views/View\\_MPH2BDamase.html](http://www.wazu.jp/gallery/views/View_MPH2BDamase.html)

### Egyptian Hieroglyphics

This paragraph shows Egyptian Hieroglyphics. These characters lie in the Unicode Supplementary Multilingual Plane U+13000 - U+1342F.



Font: Aegyptus.otf available from: <http://users.teilar.gr/~g1951d/>

SMP contains mainly ancient scripts - see <http://mPDF1.com/manual/index.php?tid=451> for full list.

mPDF uses a different method to embed fonts in the PDF file if they include characters from SMP or SIP, because the characters cannot be represented by a 4 character hex code 0000-FFFF. This method is less efficient than the default method, and it can be suppressed by adding the font name to the array 'BMPonly' in the config\_fonts.php configuration file.

Note that the DejaVu fonts distributed with mPDF and (GNU)FreeSans and FreeSerif fonts do contain a few characters in the SMP plane, but most users will not require them and by default they have been added to the array 'BMPonly'.



## USING CJK FONTS IN MPDF

Fonts containing CJK characters are large files, typically 10-30MB. Adobe provides a free download of an 'Asian font pack' allowing you to create PDF files without including (embedding) the font information in the file. This keeps the file size to a minimum and minimises resource usage on your website generating the PDF file. However, users will have to download the Adobe font packs to read the file, and other PDF software will not display the text correctly.

mPDF allows you to embed subsets of CJK fonts keeping file size down, although there is increased memory usage to generate these files.

Some CJK fonts are broken up into 2 files because of the size of the files. One freely available font with almost complete coverage of all CJK characters (in both BMP and SIP) is 'Sun' available from Alan Wood's excellent website: <http://www.alanwood.net/unicode/fonts-east-asian.html>. This comes as 2 files, Sun-ExtA and Sun-ExtB (both about 20MB in size) containing the characters from BMP and SIP respectively.

mPDF allows you to treat these as one font by defining the second file as an SIP-extension of the first in the config\_fonts.php configuration file. The following text includes random characters from the BMP and SIP mixed together:

```
□□□□□□□□□□□□□□□□□□□□□□□□□□□□
```

This is the entry in the config\_fonts.php configuration file:

```
$this->fontdata = array(
...
    "sun-exta" => array(
        'R' => "Sun-ExtA.ttf",
        'sip-ext' => 'sun-extb',
    ),
    "sun-extb" => array(
        'R' => "Sun-ExtB.ttf",
    ),
...
);
```

This is the HTML code - note only the sun-exta font-family needs to be referenced:

```
<div style="font-family:sun-extA;"> &#40706; &#40712; &#40727; &#x2320f; &#x23225; &#40742;
&#40743; &#x2322f; &#x23231; &#40761; &#40772; &#x23232; &#x23233; &#40773; &#40784; &#x23234;
&#x23256; &#40787; &#40794; &#x23262; &#x23281; &#40802; &#40809; &#x23289; &#x2328a; </div>
```

NB You may also need to edit the value \$this->useAdobeCJK=false in config.php or use new mPDF('-aCJK'), and edit the config\_cp.php configuration file.

## TRUETYPE COLLECTIONS

TrueType Collections (.ttc files) contain more than one font. mPDF treats each font separately by defining the TTCfontID array in the config\_fonts.php configuration file.

This example uses the Windows MingLiU fonts, which consist of 2 files containing 6 fonts (note that mingliub is not a Bold variant):

Font collection file (mingliu.ttc) contains the following fonts:

- [1] MingLiU (mingliu) Regular
- [2] PMingLiU (pmingliu) Regular (Proportional)
- [3] MingLiU\_HKSCS (mingliu\_hkscs) Regular

Font collection file (mingliub.ttc) contains the following fonts:

- [1] MingLiU-ExtB (mingliu-extb) Regular
- [2] PMingLiU-ExtB (pmingliu-extb) Regular (Proportional)
- [3] MingLiU\_HKSCS-ExtB (mingliu\_hkscs-extb) Regular

The following text includes characters from both BMP and SIP:

```
□ □ □ □ □ □ □ □
□ □ □ □ □ □ □ □
□ □ □ □ □ □ □ □
```

This is the entry in the config\_fonts.php configuration file:

```
$this->fontdata = array(
...
    "mingliu" => array(
        'R' => "mingliu.ttc",
        'TTCfontID' => array (
            'R' => 1,
        ),
        'sip-ext' => 'mingliu-extb',
    ),
    "pmingliu" => array(
        'R' => "mingliu.ttc",
        'TTCfontID' => array (
            'R' => 2,
        ),
        'sip-ext' => 'pmingliu-extb',
    ),
    "mingliu_hkscs" => array(
        'R' => "mingliu.ttc",
        'TTCfontID' => array (
            'R' => 3,
        ),
        'sip-ext' => 'mingliu_hkscs-extb',
    ),
    "mingliu-extb" => array(
        'R' => "mingliub.ttc",
        'TTCfontID' => array (
            'R' => 1,
        ),
    ),
    "pmingliu-extb" => array(
        'R' => "mingliub.ttc",
        'TTCfontID' => array (
            'R' => 2,
        ),
    ),
);
```

```
),
"mingliu_hkscs-extb" => array(
    'R' => "mingliub.ttc",
    'TTCfontID' => array (
        'R' => 3,
    ),
),
...
);
```

This is the HTML code:

```
<div style="font-family:mingliu;"> &#40706; &#40742; &#40772; &#40784; &#40802; &#40809;
&#x23289; &#x2328a; </div>
<div style="font-family:mingliu_hkscs;"> &#40706; &#40742; &#40772; &#40784; &#40802; &#40809;
&#x23289; &#x2328a; </div>
<div style="font-family:pmingliu;"> &#40706; &#40742; &#40772; &#40784; &#40802; &#40809;
&#x23289; &#x2328a; </div>
```