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 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

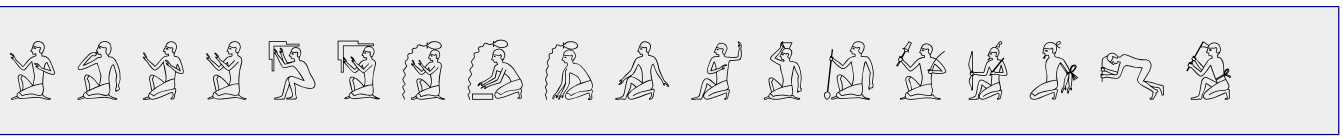
This paragraph shows Gothic text. These characters lie in the Unicode Supplementary Multilingual Plane U+10330 - U+1034F.

𐌰 𐌱 𐌲 𐌳 𐌴 𐌵 𐌶 𐌷 𐌸 𐌹 𐌺 𐌻 𐌼 𐌽 𐌾 𐌿 𐍀 𐍁 𐍂 𐍃 𐍄 𐍅 𐍆 𐍇 𐍈 𐍉 𐍊

Font: MPH2BDamase (damase_v.2.ttf) available from: 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 eficient 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'.

CJK CHARACTERS

Below are examples of all the CJK Unicode blocks contained in the Basic Multilingual Plane and Supplemental Ideographic Plane

Plane 0 (BMP)

CJK Radicals Supplement
⿰ ⿱ ⿲ ⿳ ⿴ ⿵ ⿶ ⿷ ⿸ ⿹ ⿺ ⿻

Kangxi Radicals
⼀ ⼁ ⼂ ⼃ ⼄ ⼅ ⼆ ⼇ ⼈ ⼉ ⼊ ⼋ ⼌ ⼍ ⼎ ⼏ ⼐ ⼑ ⼒ ⼓ ⼔ ⼕ ⼖ ⼗

Ideographic Description Characters
⿰ ⿱ ⿲ ⿳ ⿴ ⿵ ⿶ ⿷ ⿸ ⿹ ⿺ ⿻

CJK Symbols and Punctuation
〈 〉 《 》 「 」 『 』 【 】 〒 〓 〔 〕 〖 〗 〘 〙 〚 〛 ～ 〝 〞 〟 〠 〡 〢 〣 〤 〥 〦 〧 〨 〩 〪 〫 〬 〭 〮 〯 〰 〱 〲 〳 〴 〵 〶 〷 〸 〹 〺 〻 〼 〽 〾 〿

Hiragana
ぁ あ ぃ い ぅ う ぇ え ぉ お か が き ぎ く ぐ け げ こ ご さ ざ し じ す ず せ ぜ そ ぞ た だ ち ぢ っ つ づ て で と ど な に ぬ ね の は ば ぱ ひ び ぴ ふ ぶ ぷ へ べ ぺ ほ ぼ ぽ ま み む め も ゃ や ゅ ゆ ょ よ ら り る れ ろ ゎ わ ゐ ゑ を ん ゔ ゕ ゖ ゙ ゚ ゛ ゜ ゝ ゞ ゟ

Katakana
゠ ァ ア ィ イ ゥ ウ ェ エ ォ オ カ ガ キ ギ ク グ ケ ゲ コ ゴ サ ザ シ ジ ス ズ セ ゼ ソ ゾ タ ダ チ ヂ ッ ツ ヅ テ デ ト ド ナ ニ ヌ ネ ノ ハ バ パ ヒ ビ ピ フ ブ プ ヘ ベ ペ ホ ボ ポ マ ミ ム メ モ ャ ヤ ュ ユ ョ ヨ ラ リ ル レ ロ ヮ ワ ヰ ヱ ヲ ン ヴ ヵ ヶ ヷ ヸ ヹ ヺ ・ ー ヽ ヾ ヿ

Bopomofo
ㄅ ㄆ ㄇ ㄈ ㄉ ㄊ ㄋ ㄌ ㄍ ㄎ ㄏ ㄐ ㄑ ㄒ ㄓ ㄔ ㄕ ㄖ

Hangul Compatibility Jamo
ㄱ ㄲ ㄳ ㄴ ㄵ ㄶ ㄷ ㄸ ㄹ ㄺ ㄻ ㄼ ㄽ ㄾ ㄿ ㅀ ㅁ ㅂ

Kanbun
㆐ ㆑ ㆒ ㆓ ㆔ ㆕ ㆖ ㆗ ㆘ ㆙ ㆚ ㆛ ㆜ ㆝ ㆞ ㆟

Bopomofo Extended
ㄠ ㄡ ㄢ ㄣ ㄤ ㄥ ㄦ ㄧ ㄨ ㄩ ㄪ ㄫ ㄬ ㄭ ㄮ ㄯ ㄰ ㄱ ㄲ ㄳ ㄴ

Katakana Phonetic Extensions
ㇰ ㇱ ㇲ ㇳ ㇴ ㇵ ㇶ ㇷ ㇸ ㇹ ㇺ ㇻ ㇼ ㇽ ㇾ ㇿ

Enclosed CJK Letters and Months
㈠ ㈡ ㈢ ㈣ ㈤ ㈥ ㈦ ㈧ ㈨ ㈩ ㈪ ㈫ ㈬ ㈭ ㈮ ㈯ ㈰ ㈱ ㈲ ㈳ ㈴ ㈵ ㈶ ㈷ ㈸ ㈹ ㈺ ㈻ ㈼ ㈽ ㈾ ㈿ ㉀ ㉁ ㉂ ㉃ ㉄ ㉅ ㉆ ㉇ ㉈ ㉉ ㉊ ㉋ ㉌ ㉍ ㉎ ㉏

CJK Compatibility
㌀ ㌁ ㌂ ㌃ ㌄ ㌅ ㌆ ㌇ ㌈ ㌉ ㌊ ㌋ ㌌ ㌍ ㌎ ㌏ ㌐ ㌑ ㌒ ㌓ ㌔ ㌕ ㌖ ㌗ ㌘ ㌙ ㌚ ㌛ ㌜ ㌝ ㌞ ㌟ ㌠ ㌡ ㌢ ㌣ ㌤ ㌥ ㌦ ㌧ ㌨ ㌩ ㌪ ㌫ ㌬ ㌭ ㌮ ㌯

CJK Unified Ideographs Extension A
㐀 㐁 㐂 㐃 㐄 㐅 㐆 㐇 㐈 㐉 㐊 㐋 㐌 㐍 㐎 㐏 㐐 㐑 㐒 㐓 㐔 㐕 㐖 㐗 㐘 㐙 㐚 㐛 㐜 㐝 㐞 㐟 㐠 㐡 㐢 㐣 㐤 㐥 㐦 㐧 㐨 㐩 㐪 㐫 㐬 㐭 㐮 㐯

CJK Unified Ideographs
一 丁 丂 七 丄 丅 丆 万 丈 三 上 下 丌 不 与 丏 丐 丑 丒 专 且 丕 世 丗 丘

Yi Syllables
ꀀ ꀁ ꀂ ꀃ ꀄ ꀅ ꀆ ꀇ ꀈ ꀉ ꀊ ꀋ ꀌ ꀍ ꀎ ꀏ ꀐ ꀑ ꀒ ꀓ ꀔ ꀕ ꀖ ꀗ ꀘ ꀙ ꀚ ꀛ ꀜ ꀝ ꀞ ꀟ ꀠ ꀡ ꀢ ꀣ ꀤ ꀥ ꀦ ꀧ ꀨ ꀩ ꀪ ꀫ ꀬ ꀭ ꀮ ꀯ

Yi Radicals
ꀰ ꀱ ꀲ ꀳ ꀴ ꀵ ꀶ ꀷ ꀸ ꀹ ꀺ ꀻ ꀼ ꀽ ꀾ ꀿ ꁀ ꁁ ꁂ ꁃ ꁄ ꁅ ꁆ ꁇ ꁈ ꁉ ꁊ ꁋ ꁌ ꁍ ꁎ ꁏ

Hangul Syllables
가 각 갂 갃 간 갅 갆 갇 갈 갉 갊 갋 갌 갍 갎 갏 감 갑 값 갓 갔 강 갖 갗 갘

CJK Compatibility Ideographs
豈 更 車 賈 滑 串 句 龜 契 金 喇 奈 懶 癩 羅 蘿 螺 裸 邏 樂 洛 烙 珞 落 酪

Plane 2 (SIP)

CJK Unified Ideographs Extension B
𠀀 𠀁 𠀂 𠀃 𠀄 𠀅 𠀆 𠀇 𠀈 𠀉 𠀊 𠀋 𠀌 𠀍 𠀎 𠀏 𠀐 𠀑 𠀒 𠀓 𠀔 𠀕 𠀖 𠀗 𠀘 𠀙 𠀚 𠀛 𠀜 𠀝 𠀞 𠀟 𠀠 𠀡 𠀢 𠀣 𠀤 𠀥 𠀦 𠀧 𠀨 𠀩 𠀪 𠀫 𠀬 𠀭

CJK Unified Ideographs Extension C
𪜀 𪜁 𪜂 𪜃 𪜄 𪜅 𪜆 𪜇 𪜈 𪜉 𪜊 𪜋 𪜌 𪜍 𪜎 𪜏 𪜐 𪜑 𪜒 𪜓 𪜔 𪜕 𪜖 𪜗 𪜘 𪜙 𪜚 𪜛 𪜜 𪜝 𪜞 𪜟 𪜠 𪜡 𪜢 𪜣 𪜤 𪜥 𪜦 𪜧 𪜨 𪜩 𪜪 𪜫 𪜬 𪜭

CJK Compatibility Ideographs Supplement
丽 丸 乁 亘 你 侮 悅 倂 僧 備 僧 像 儁 充 兔 兎 爐 具 與 巔

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:

```
𪜀 𪜁 𪜂 𪜃 𪜄 𪜅 𪜆 𪜇 𪜈 𪜉 𪜊 𪜋 𪜌 𪜍 𪜎 𪜏 𪜐 𪜑 𪜒 𪜓 𪜔 𪜕 𪜖 𪜗 𪜘 𪜙
```

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

This is a sample 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 sample 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>
```