WIKIPEDIA

GNU Unifont

The **GNU Unifont** by Roman Czyborra is a free Unicode bitmap font using an intermediate bitmapped font format. The main Unifont covers the entire Basic Multilingual Plane (BMP), the "Upper" companion covers significant parts of the Supplementary Multilingual Plane, and the "Unifont JP" companion contains Japanese kanji present in the JIS X 0213 character set.

It is present in most free operating systems and windowing systems such as Linux, XFree86 or the X.Org Server and some embedded firmware such as RockBox. The font is released under the GNU General Public License Version 2+ with a font embedding exception (embedding the font in a document does not require the document to be placed under the same license).

It became a GNU package in October 2013. The current maintainer is Paul Hardy.

Contents

Status

The Unicode Basic Multilingual Plane covers 2^{16} = 65,536 code points. Of this number, 2048 are reserved for special use as UTF-16 surrogate pairs and 6,400 are reserved for private use. This leaves 57,088 code points to which glyphs can be assigned. Some of these code points are special values that do not have an assigned glyph, but most do have assigned glyphs.

As of May 2019, the GNU Unifont has complete coverage of the Basic Multilingual Plane as defined in Unicode 12.1.0. Its companion fonts, Unifont Upper and Unifont CSUR, have significant coverage of the Supplementary Multilingual Plane and the ConScript Unicode Registry, respectively.

GNU Unifont

Category	Unicode, Bitmap, Sans-serif
Classification	Duospace
Designer(s)	Roman Czyborra, Paul Hardy
Date created	1998
Date released	12.1.03 (August 11, 2019) [±] (https://en.wikipedia.org/w/index.php?title=Template:Latest_stable_software_release/GNU_Unifont&action=edit)
License	GNU General Public License
See all characters	
Website	unifoundry.com/unifont/index.html (http://unifoundry.com/unifont/index.html)

For version 12.1.02, Unifont JP was released, which covers 10,000 Japanese kanji present in the JIS X 0213 character set, some of which are in the Supplementary Ideographic Plane. It is derived from Jiskan16, a public domain font.

Scripts that are less than 100% complete can be augmented by any contributor.

The large block of about 20,000 CJK ideographs has been copied from WenQuanYi's Unibit font with permission.[1](Wen Quan Yi: Spring of Letters)

However, despite of its coverage, Unifont stores only one glyph per printable Unicode code point. So it does not feature the OpenType features needed to render correctly scripts with complex layouts (notably Indic scripts), and correctly position the combining diacritics with base letters if these combinations are not encoded in Unicode in their precombined form; as well the contextual forms (including joining types, and subjoined clusters) are not handled: this would increase the number of glyphs to include in the basic font and it's still not possible (because of current OpenType limitations) to encode all the needed glyphs to represent all the required combinations that can exist in a single Unicode plane (this is also true for Chinese fonts that cannot cover completely all ideograms currently encoded in two planes, and in also in a third planes starting with the next version 13.0 of Unicode). Such font can then only be used as a "last resort" default font, suitable for simple alphabetic scripts, or to render isolated characters, but will make actual texts difficult or sometimes impossible to read correctly. For correctly rendering Indic abugidas (and Semitic abjads if they are written with their optional combining diacritics), other fonts should be specified in stylesheets before this one, and additional fonts will be needed to cover Han ideographs encoded in supplementary planes, or to render most historic (or minority modern) scripts not encoded in the BMP.

Distribution

Unifont, as of version 12.0.0, is available in vector TTF, BDF, and PCF formats for the "standard build". Only the TrueType build is split into Unifont and two companion fonts.[1]

A few "specialized versions" have been built by request and made available by Paul Hardy. These include a bitmap TTF (SBIT) with empty glyphs filled with code-point values for FontForge users to read, a PSF bitmap with glyphs for APL programmers, and single-file versions in Roman's .hex format (see below).[1] The actual organization of the source consists of smaller .hex files to be stitched together and converted to other formats in a build.[2]

Sample in Japanese and Chinese

Vectorization

Luis Alejandro González Miranda wrote scripts to vectorize and convert the BDF font to TrueType format using FontForge.[3] Paul Hardy adjusted these scripts to handle combining characters (accents, etc.) for the latest TrueType versions.[1](TrueType Font Generation)

The .hex font format

The GNU Unifont .hex format defines its glyphs as either 8 or 16 pixels in width by 16 pixels in height. Most Western script glyphs can be defined as 8 pixels wide, while other glyphs (notably the Chinese-Japanese-Korean, or CJK set) are typically defined as 16 pixels wide.

The unifont.hex file contains one line for each glyph. Each line consists of a four digit Unicode hexadecimal code point, a colon, and the bitmap string. The bit string is 32 hexadecimal digits for an 8 pixel wide glyph or 64 hexadecimal digits for a 16 pixel wide glyph. The goal is to create an intermediate format that would facilitate adding new glyphs.

A '1' bit in the bit string corresponds to an 'on' pixel. Pixels bits are stored top to bottom, left to right.

Example

This is an example font containing one glyph, for ASCII capital 'A'.

```
0041:00000000182424244 27E424242420000
```

The first number is the hexadecimal Unicode code point, with range 0000 through FFFF. Hexadecimal 0041 is decimal 65, the code point for the letter 'A'. The colon separates the code point from the bitmap. In this example, the glyph is 8 pixels wide, so the bit string is 32 hexadecimal digits long.

The bit string begins with 8 zeros, so the top 4 rows will be empty (2 hexadecimal digits per 8 bit byte, with 8 bits per row for an 8 pixel-wide glyph). The bit string also ends with 4 zeros, so the bottom 2 rows will be empty. It is implicit from this that the default font descender is 2 rows below the baseline, and the capital height is 10 rows above the baseline. This is the case in the GNU Unifont with Latin glyphs.

Over time a number of ways have been derived to handle the format. The earliest way is the `hexdraw` Perl script, which converts the string into an ASCII art representation to be edited in a text editor. As of April 2019, the current way involves generating a bitmap image grid for an entire range of code points and working with an image editor. In either case, the edited glyphs are converted back into .hex files for storage.[2]

Hexdraw representation of the example

Actual output	Spaced out for ease-of-reading

```
0041:                          0041:
     --------                       - - - - - - - -
     --------                       - - - - - - - -
     --------                       - - - - - - - -
     --------                       - - - - - - - -
     ---##---                       - - - # # - - -
     --#--#--                       - - # - - # - -
     --#--#--                       - - # - - # - -
     -#----#-                       - # - - - - # -
     -#----#-                       - # - - - - # -
     -######-                       - # # # # # # -
     -#----#-                       - # - - - - # -
     -#----#-                       - # - - - - # -
     -#----#-                       - # - - - - # -
     -#----#-                       - # - - - - # -
     --------                       - - - - - - - -
     --------                       - - - - - - - -
```

History

Roman Czyborra created the Unifont format in 1998[4] after earlier efforts dating to 1994.

In 2008, Luis Alejandro González Miranda wrote a program to convert this font into a TrueType font. Paul Hardy modified it later to support combining characters in the TrueType version.

Finally, Richard Stallman dubbed Unifont a GNU package in October 2013, with Paul Hardy as its maintainer.

References

1. *GNU Unifont Glyphs* (http://unifoundry.com/unifont.html)
2. "Unifoundry Unicode Utilities" (http://unifoundry.com/unifont/unifont-utilities.html). *unifoundry.com*. Retrieved 16 April 2019.

3. *GNU Unifont in TrueType format* (https://web.archive.org/web/20160201193830/http://www.lgm.cl/trabajos/unifont/index.en.html), archived from the original (http://www.lgm.cl/trabajos/unifont/index.en.html) on 2016-02-01

4. Roman Czyborra's GNU Unifont page (http://czyborra.com/unifont/)

- The Unicode Consortium: *The Unicode 5.0 Standard*. 5th, Addison Wesley 2007; ISBN 0-321-48091-0.

External links

- GNU Project Archives (https://ftp.gnu.org/gnu/unifont/)
- Unifoundry.com GNU Unifont page (http://unifoundry.com/unifont.html)

Retrieved from "https://en.wikipedia.org/w/index.php?title=GNU_Unifont&oldid=916948745"