Typographic Names in Unicode CLDR

Sascha Brawer (Google) and Sairus Patel (Adobe)
December 27, 2017

Objective

We'd like to store typographic terminology in the Unicode Common Locale Data Repository (CLDR). This is useful for apps that show the name of font styles and design axes according to the user's languages. It would also be useful for system-level libraries such as Apple's CoreText or Microsoft's DirectWrite.

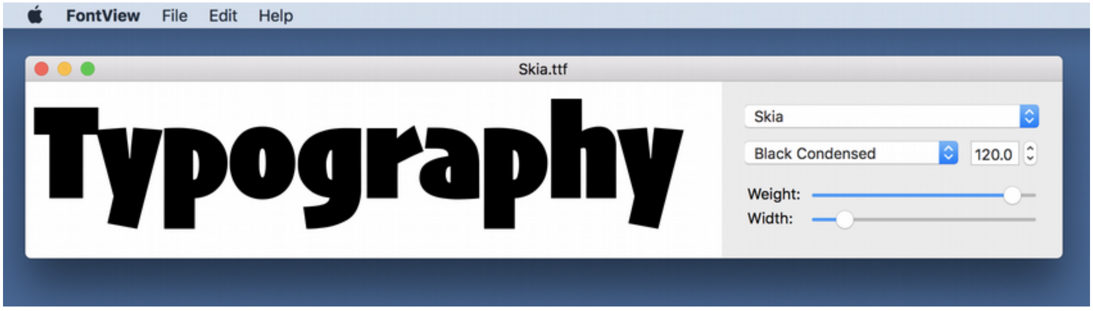

Fig. 1: FontView, a simple open-source font viewer. The strings in the right-hand panel ("Black," "Condensed," "Weight," "Width") are taken from the font file. OpenType and TrueType support localized names, but most fonts only contain English strings. Even when a font does have translations, coverage is typically limited to a very small number of languages. If typographic terminology were part of Unicode CLDR, applications and system libraries could supplement the font data with localized names from CLDR.

```
CFStringRef nameLanguage = NULL;
CFStringRef name = CTFontCopyLocalizedName(font, kCTFontStyleNameKey, &nameLanguage);
```

Fig. 2: Apple's CoreText library allows applications to retrieve the name for a font style in the user's language. For example, this function might return "Extrafett Schmal" when the user speaks German. Other systems offer similar functions.

Axis and style names

The Unicode CLDR data is maintained in XML format. For axis and style names, we propose this:

```
<!-- In http://unicode.org/repos/cldr/trunk/common/main/en.xml -->
<typographicNames>
    <axisName type="wght">Weight</axisName>
    <axisName type="wdth">Width</axisName>
    <styleName type="wght" subtype="100">Thin</styleName>
    <styleName type="wght" subtype="200">Ultralight</styleName>
    <styleName type="wght" subtype="300">Light</styleName>
    <styleName type="wght" subtype="400">Normal</styleName>
    <styleName type="wght" subtype="500">Medium</styleName>
    <styleName type="wght" subtype="600">Semibold</styleName>
    <styleName type="wght" subtype="700">Bold</styleName>
    <styleName type="wght" subtype="800">Extrabold</styleName>
    <styleName type="wght" subtype="900">Black</styleName>
    <styleName type="wdth" subtype="50">Ultracondensed</styleName>
    <styleName type="wdth" subtype="75">Condensed</styleName>
    <styleName type="wdth" subtype="100">Regular</styleName>
</typographicNames>
```

The XML structure and attribute names are in line with the general setup of CLDR, so they can be supported by Unicode's existing toolchain without major changes. The axis identifiers ("wght", "wdth") and the numeric values ("100", "200", etc.) are taken from the OpenType Design-Variation Axis Tag Registry.

Translations

Translations will use the same format as the English data:

```
<!-- In http://unicode.org/repos/cldr/trunk/common/main/de.xml -->
<typographicNames>
    <axisName type="wght">Stärke</axisName>
    <styleName type="wght" subtype="100">Mager</styleName>
    <styleName type="wght" subtype="200">Ultraleicht</styleName>
</typographicNames>
```

Alternate names

CLDR has a mechanism for alternate names, for example to store "Czechia" and "Czech Republic" as alternate names for the same country. The same works for typographic names:

```
<typographicNames>
    <styleName type="wght" subtype="200">Ultralight</styleName>
    <styleName type="wght" subtype="200" alt="variant">Extralight</styleName>
</typographicNames>
```

Slanted and Backslanted

Other than for the named styles along the *Width, Weight,* and *Italic* styles, there are no clearly defined values for the styles "Slanted" and "Backslanted" in the OpenType Design-Variation Axis Tag Registry. To identify these styles in CLDR, we suggest the following. Fonts are *not* expected to use a particular slant angle in their designs; we simply need some identifier.

```
<-- In http://unicode.org/repos/cldr/trunk/common/main/en.xml -->
<typographicNames>
    <axisName type="ital">Italic</axisName>
    <styleName type="ital" subtype="12">Italic</styleName>
    <axisName type="slnt">Slant</axisName>
    <styleName type="slnt" subtype="12">Oblique</styleName>
    <styleName type="slnt" subtype="-12" alt="variant">Backslanted</styleName>
</typographicNames>
```

Feature names

For typographic features, we propose the following representation:

```
<typographicNames>
    <featureName type="smcp">Small capitals</featureName>
    <featureName type="smcp" alt="short">Smallcaps</featureName>
    <featureName type="zero">Slashed zero</featureName>
</typographicNames>
```

As identifiers, we'll use the tags from the OpenType Feature Tag Registry. Given their large number, we plan to only put the names of frequently-used OpenType feature names into CLDR. (Many features are not user-visible settings, but instead serve as a data channel for software to pass information to the font).

Tests for data integrity

CLDR runs automated checks on the translations. For example, this would flag a Chinese character inside a French string. Beyond the existing tests which CLDR runs on all its data files, we'll implement a new check to make sure that the style names for a given design axis are unique in every language.

Seeding the translations

For seeding CLDR, Apple, Microsoft and Adobe have contributed their translations.

Appendix: Generating localized subfamily names

To find a localized subfamily name such as "Extraleicht Schmal" for a font called "Extralight Condensed", a system or application library might do the following:

1. Determine the set of languages in which the subfamily name can potentially be returned. This is the union of the languages for which the font contains 'name' table entries with ID 2 or 17, plus the languages for which CLDR supplies typographic names.
2. Run a BCP47 language matching algorithm to find the best available language given the user preferences. The resulting subfamily name will be localized to this language.
3. If the font's 'name' table contains a *typographic subfamily name* (ID 17) in this language and all font variation axes are set to their defaults, return this name.
4. If the font's 'name' table contains a *font subfamily name* ('name' ID 2) in this language and all font variation axes are set to their defaults, return this name.
5. If the font has a style attributes (STAT) table, look up the design axis tags and their ordering. If the font has no STAT table, assume [Width, Weight, Slant] as axis ordering, and infer the font's style attributes from other available data in the font (eg. the OS/2 table).
6. For each design axis, find a localized style name for its value.
   - 6.1 If the font's style attributes point to a 'name' table entry that is available the result language, use this name.
   - 6.2 Otherwise, generate a fallback name from CLDR *styleName* data.
     - 6.2.1 The *type* key is the OpenType axis tag ('wght'). The *subtype* and *alt* keys are taken from the entry in English CLDR where the string is equal to the English name in the font. For example, when the font uses a weight whose English style name is "Extralight", this will lead to *subtype* = "200" and *alt* = "variant". If there is no match, take the axis value ("200") for *subtype* and the empty string for *alt*.
     - 6.2.2 Look up (*type, subtype*) in a data table derived from CLDR's style names. If CLDR supplies multiple alternate names for this (type, subtype), use the one whose "alt" key is matching; otherwise, use the default alternate (which has no "alt" attribute in CLDR).
7. Concatenate the strings from step 6 according to CLDR ticket #10 752.

The above is justone potential approach for using the CLDR data. Of course, applications are free to implement their own algorithms depending on their specific needs.