East Asian typography on the modern web

By Chen Hui Jing / @hj_chen
Surname: Chen
First name: Hui Jing
@hj_chen
Overview

一、Recognising the characters
二、Displaying the characters
三、Laying out the characters
Character encoding

<table>
<thead>
<tr>
<th>Encoding options available in FontForge</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO 8859-1 (Latin1)</td>
</tr>
<tr>
<td>ISO 8859-13 (Latvian)</td>
</tr>
<tr>
<td>ISO 8859-2 (Latin2)</td>
</tr>
<tr>
<td>ISO 8859-4 (Latin4)</td>
</tr>
<tr>
<td>ISO 8859-9 (Latin5)</td>
</tr>
<tr>
<td>ISO 8859-15 (Latvian)</td>
</tr>
<tr>
<td>ISO 8859-16 (Latin7)</td>
</tr>
<tr>
<td>ISO 8859-19 (Latin11)</td>
</tr>
<tr>
<td>ISO 8859-8 (Cyrillic)</td>
</tr>
<tr>
<td>KOI8-R (Cyrillic)</td>
</tr>
<tr>
<td>ISO 8859-3 (Arabic)</td>
</tr>
<tr>
<td>ISO 8859-7 (Hebrew)</td>
</tr>
<tr>
<td>ISO 8859-11 (Thai)</td>
</tr>
<tr>
<td>Macintosh Latin</td>
</tr>
<tr>
<td>Windows Latin &quot;ANSI&quot;</td>
</tr>
<tr>
<td>AKAIDC Standard</td>
</tr>
<tr>
<td>TCGA Base (BR)</td>
</tr>
<tr>
<td>ISO 10646-1 (Unicode, BMP)</td>
</tr>
<tr>
<td>ISO 10646-1 (Unicode, UTF8)</td>
</tr>
<tr>
<td>Shift (Kurd)</td>
</tr>
<tr>
<td>JIS 201 (Kanji)</td>
</tr>
<tr>
<td>JIS 212 (Kanji)</td>
</tr>
<tr>
<td>Kanji (Korean)</td>
</tr>
<tr>
<td>KOI8-M (Korean)</td>
</tr>
<tr>
<td>JIS2001 (Korean)</td>
</tr>
<tr>
<td>JIS2012 (Chinese, Simplified)</td>
</tr>
<tr>
<td>GB2312 (Chinese)</td>
</tr>
<tr>
<td>Big5 (Traditional Chinese)</td>
</tr>
<tr>
<td>Big5-HKSCS (Traditional Chinese)</td>
</tr>
</tbody>
</table>
Declaring character encodings for the web

```html
<!doctype html>
<html lang="en">
<head>
<meta charset="utf-8"/>
...

@charset "utf-8";
/* all the styles for your web page */
```
Text rendering on electronic screens
Bitmap fonts

How razorback-jumping frogs can level six piqued gymnasts!

A letter as designed, pixels, run-length coding.

https://huijing.github.io/slides/81-typetechmunich-2020
The Font Wars by Typography.Guru

https://youtu.be/5X9Dj7tBlkg
## Font formats for web use

<table>
<thead>
<tr>
<th>Format</th>
<th>Microsoft</th>
<th>Other Libraries</th>
</tr>
</thead>
<tbody>
<tr>
<td>WOFF2 (Web Open Font Format 2)</td>
<td><img src="#" alt="Microsoft" /></td>
<td><img src="#" alt="Mozilla" /></td>
</tr>
<tr>
<td>WOFF (Web Open Font Format)</td>
<td><img src="#" alt="Microsoft" /></td>
<td><img src="#" alt="Mozilla" /></td>
</tr>
<tr>
<td>OTF (OpenType)</td>
<td><img src="#" alt="Microsoft" /></td>
<td><img src="#" alt="Apple" /></td>
</tr>
<tr>
<td>TTF (TrueType)</td>
<td><img src="#" alt="Microsoft" /></td>
<td><img src="#" alt="Apple" /></td>
</tr>
<tr>
<td>EOT (Embedded Open Type)</td>
<td><img src="#" alt="Microsoft" /></td>
<td></td>
</tr>
<tr>
<td>SVG (Scalable Vector Graphics)</td>
<td><img src="#" alt="W3C SVG" /></td>
<td></td>
</tr>
</tbody>
</table>

https://huijing.github.io/slides/81-typetechmunich-2020
Chinese typeface design
3. Adding WebFonts to CSS

The font description provides the bridge between a font reference (within a stylesheet) and the font data, which is the data needed to format text and to render the glyph representations to which the characters map—the actual scalable outlines or bitmaps. Fonts are referenced by style sheet properties. The font description is used to select the relevant font data. The font description consists of descriptors which provide the location of the font data on the Web, and/or characterize that font data. The font descriptors are also used to match font references with the correct font description.

Font descriptors may be classified into three types:

1. those that provide the link between the CSS usage of the font and the font description—these have the same names as the corresponding CSS font properties
2. the URL for the location of the font data
3. those that further characterize the font, to provide a link between the font description and the font data

Font descriptions are contained in an @-rule. At-rules—so called because they start with the @ symbol—were introduced in CSS 1 where they were used to import other style sheets. Ordinary CSS rules start with a selector, and apply solely to the selected HTML elements. At-rules have no selector, and apply to the whole style sheet.

The general form of this new at-rule is:

```
@font-face { <font description> }
```

where the font description has the form

```javascript
descriptor: value;
descriptor: value;
[...] descriptor: value;
```

For example, here the font 'Robson Celtic' is defined and referenced in a style sheet contained in an HTML document.

```
@font-face {
  font-family: robson-celtic;
  src: url('robson-celtic-webfont.woff2') format('woff2'),
       url('robson-celtic-webfont.woff') format('woff');
}
```

Earliest reference of webfonts I could find (21-Jul-1997)
Font stacks

/* This text is in Lucida Grande */
.sans {
}

/* A typical Chinese font stack, declare Latin fonts first */
.zh-hans {
  font-family: Tahoma, Helvetica, Arial, "Microsoft Yahei", "微软雅黑", STXihei, "华文细黑", sans-serif;
}
Subset fail
```css
@font-face {
  font-family: 'Bellato';
  src: url('Bellota-Regular.woff2') format('woff2'),
       url('Bellota-Regular.woff') format('woff');
}
```
Anatomy of an @font-face rule

@font-face {
  font-family: <family-name>;
  src: [ <url> [format(<string> #)]? | <font-face-name> ] #;]
  font-style: normal | italic | oblique ;
  font-weight: normal | bold | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900;
  font-stretch: normal | ultra-condensed | extra-condensed | condensed | semi-condensed | semi-expanded;
  unicode-range: <range> #;
  font-variant: normal | none | [ <common-lig-values> | <discretionary-lig-values> | <historical-lig-values> ] #;
  font-feature-settings: normal | <feature-tag-value> #;
}
The backdrop is the content behind the element and is what the element is composited with.

Slides for *This world mixed and blended* by Gao Wei.
Raleway
Multiple Designers (18 styles)

This font does not support Cyrillic characters

Comfortaa
Johan Aakerlund, Cyreal (5 styles)

Этот шрифт поддерживает символы кириллицы
@font-face {
  font-family: 'Raleway';
  src: url('fonts/raleway-regular.woff2') format('woff2'),
       url('fonts/raleway-regular.woff') format('woff');
  /* no range specified, defaults to entire range */
}

@font-face {
  font-family: 'Raleway';
  src: url('fonts/comfortaa_regular.woff2') format('woff2'),
       url('fonts/comfortaa_regular.woff') format('woff');
  unicode-range: U+0400-U+04FF; /* Unicode range for Cyrillic characters */
}
The Russian word for "Thank You" is "Спасибо".
Source Hans Serif (65,535 glyphs)
Subsetting web fonts

Font Squirrel

FontPlop

Font Spider

https://huijing.github.io/slides/81-typetechmunich-2020
Variable fonts

DRAGONS

Space

Weight
Progressive Font Enrichment

“to enable the ability for only the required part of the font be downloaded on any given page, and for subsequent requests for that font to dynamically ‘patch’ the original download with additional sets of glyphs as required on successive page views—even if they occur on separate sites

—Jason Pamental on Web Fonts & Typography News #11
Incremental Transfer Demo

Transfer Options

NOTE: all options other than incremental transfer break layout features across segments. In reality this forces the use of large blocks in many scenarios (Arabic, Indic, many-fatin cases, etc).
148 codepoints (17 + 21 + 23 + 64 + 23) in Demo Content

Options:
A) What Google Fonts would send today, 2 size 34.9 KB:
[13.8 KB cursive] [21.2 KB latin]
B) wof2 of each segment, 2 segments 21.7 KB:
[4.3 KB wof2] [5.5 KB wof2] [6.1 KB wof2] [3.0 KB wof2] [2.8 KB wof2]
C) Incremental Transfer, 2 patches 10.3 KB:
[4.9 KB patch] [2.9 KB patch] [2.8 KB patch]
D) Optimal, wof2 of the exact subset:
[8.8 KB wof2]

Demo Content

Google Fonts’ Incremental Transfer Demo
Hi, I’m Jason

My name is Jason Pamental. I’m a design strategist, UX leader, technologist, expert in web typography, and Invited Expert on the W3C Web Fonts Working Group. I’ve spoken with organizations like Adobe, Audible, Condé Nast, GoDaddy, IBM—and given presentations and workshops at conferences all over the world. I also publish a newsletter on web typography.

Type is how we ‘hear’ what we read

Since the introduction of Variable Fonts in late 2016, I’ve spent a good deal of time researching, writing about, and working with them, and am convinced they are the future of type—and indeed design—on the web. I’ve designed and published a Type Network and a responsive web typography with an article and demo page.
Typesetting on the web

“Authors should language-tag their content accurately for the best typographic behaviour.”

—CSS Text Module Level 3
OpenType features
Font feature properties

- **font-kerning**
  - No kerning
  - Kerning applied

- **font-variant-position**
  - C H N O P
  - C H N O P

- **font-variant-position-ligatures**
  - fi > fi
  - tz > t
  - Words > Words
  - labor of love > labor of love

- **font-variant-numeric**
  - 1st 17th 2a ▶ 1st 17th 2a
  - Lining Old-Style
  - 409,280 367,112 155,068 171,792
  - 409,280 367,112 155,068 171,792

- **font-variant-caps**
  - The DOM, the HTML syntax, and the XHTML syntax cannot all represent the same content. For example, namespaces cannot be represented using the HTML syntax, but they are supported in the DOM and in the XHTML syntax.

- **font-variant-alternates**
  - Jesuits ▶ Jefuits
  - quick ▶ quick
  - Quick ▶ Quick
  - incroyable ▶ incroyables
  - 519 ▶ 519

[Link to original document](https://huijing.github.io/slides/81-typetechmunich-2020)
Font metrics are crucial
font-variant-east-asian

Allows control of glyph substitution and sizing in East Asian text

一個簡體字可能對應多個繁體字，如簡體字「發」，其相應的繁體字可能為「發」或「髮」；一個繁體漢字對應多個簡體漢字的情況與前者相比數量極少但仍需註意，如繁體字「乾」可能對應簡體字「幹」或「乾」。繁簡漢字的對應關係具體應由上下文決定。

茨 噓
`font-language-override`

To control the use of language-specific glyph substitutions and positioning

Example lifted from CSS Fonts Module Level 4

```html
e--- Macedonian lang code --
<body lang="mk">
  <h4>Член 9</h4>
  <p>Никој човек нема да биде подложен на произволно апсене, притвор или прогонување.</p>
</body>

body {
  /* Serbian OpenType language tag */
  font-language-override: "SRB";
}
```
The text-transform property

If I want [flowers], I'm going to send them to myself.

Süßes Sofa-Klößchen genießen maßgeblich gefräßige preußische Nutznicher.

Οὐδέν κακόν αμιγές καλοῦ.

アイウェオカクケ
The text-emphasis property

We use *italics* to emphasise words in English, 但是中文则是用着重号.

```html
<p>We use <em>italics</em> to emphasise words in E
```
text-align & text-justify

- Sets the text-align-all and text-align-last properties
- Describes how the inline-level content of a block is aligned along the inline axis if the content does not completely fill the line box.

```
text-align: start | end | left | right | center | justify | match-parent | justify-all
```

Values other than justify-all or match-parent are assigned to text-align-all and reset text-align-last to auto.

Selects the justification method used when a line's alignment is set to justify

```
text-justify: auto | none | inter-word | inter-character
```
As you can see, the text is aligned to the left margin meaning that not all of the text will align to the right margin. Here’s an example of justified website text. As you can see, the text is aligned to both the left margin and right margins, meaning that each line will be the same length.

Don’t use fully justified text alignment on your website

People also ask

Should I justify text?

How do I manually justify text?

How do you justify text in CSS?

What is causing the text to be justified?
Text alignment and justification

定义了CSS如何支持各种不同国际化语言的书写模式，例如拉丁（Latin）语系及印度（Indic）语系采用从左到右的书写模式，希伯来语（Hebrew）或阿拉伯语（Arabic）采用从右到左的书写模式，一些混合了拉丁语和阿拉伯语的文字可能采用双向书写（bidirectional），而一些东亚文字则需要竖排（从上到下）的书写模式。
Vertical text on the web

CSS Writing Modes Level 3 defines CSS features to support for various international writing modes, such as left-to-right (e.g. Latin or Indic), right-to-left (e.g. Hebrew or Arabic), bidirectional (e.g. mixed Latin and Arabic) and vertical (e.g. Asian scripts).
**writing-mode property**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>horizontal-tb</td>
<td>从1987到现在</td>
</tr>
<tr>
<td>vertical-rl</td>
<td>从1987到现在</td>
</tr>
<tr>
<td>vertical-lr</td>
<td>从1987到现在</td>
</tr>
<tr>
<td>sideways-rl*</td>
<td>从1987到现在</td>
</tr>
<tr>
<td>sideways-lr*</td>
<td>从1987到现在</td>
</tr>
</tbody>
</table>

Properties marked with * have been deferred to Writing Modes Level 4.
text-orientation property

从1987到现在
mixed

从1987到现在
upright

从1987到现在
sideways
text-combine-upright property

none

all
digits <integer>*

Codepen example for text-combine-upright

https://huijing.github.io/slides/81-typetechmunich-2020
古巴
Cuba

乌克兰
Ukraine

列支敦士登
Liechtenstein
Not just for East Asian text

Das Leben ist kein Ponyhof

Es liegt der heiße Sommer
Auf deinen Wangelein;  
Es liegt der Winter, der kalte, 
In deinem Herzchen klein.

Das wird sich bei dir ändern,
Du Vielgeliebte mein!  
Der Winter wird auf den Wangen,  
Der Sommer im Herzen sein.
Everything is a box on the web
Line breaks in inline boxes

If an element generates zero boxes, was it really there at all?

```html
<p class="line-container">If an element <em>generates zero boxes</em>, was it <strong>really there</strong> at all?</p>

.linebox p.line-container {
  font-size: 150%;
}

.linebox .line-container em {
  background-color: limegreen;
  padding: 0.25em;
  mix-blend-mode: color;
}
```
## CSS for controlling line breaks

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>line-break</td>
<td>allows choosing various levels of “strictness” for line breaking restrictions</td>
</tr>
<tr>
<td>word-break</td>
<td>controls what types of letters are glommed together to form unbreakable “words”, causing CJK characters to behave like non-CJK text or vice versa</td>
</tr>
<tr>
<td>hyphens</td>
<td>controls whether automatic hyphenation is allowed to break words in scripts that hyphenate</td>
</tr>
<tr>
<td>overflow-wrap</td>
<td>allows the UA to take a break anywhere in otherwise-unbreakable strings that would otherwise overflow</td>
</tr>
</tbody>
</table>

*Line breaking* by Florian Rivoal @ dotCSS
Future improvements to web typography

- CSS Rhythmic Sizing
- CSS Inline Layout Module Level 3
- CSS Text Module Level 4
- CSS Text Decoration Module Level 4
The leading problem
The interruption problem

Neque porro quisquam est, qui dolorem ipsum quia

Neque porro quisquam est, qui dolorem ipsum quia dolor sit amet, consectetur, adipisicing elit, sed quia non numquam eius modi tempora incidunt ut labore et dolore magna aliqua quia quaeat voluptatem.

Ut enim ad minim veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur? Quis autem vel eum iure reprehenderit qui in ea voluptate velit
CSS Line Layout and Vertical Rhythm
Imagine an alternate history...
References

- The Absolute Minimum Every Software Developer Absolutely, Positively Must Know About Unicode and Character Sets (No Excuses!)
- Choosing & applying a character encoding
- Fonts and Keyboards
- Type rendering on the web
- Mac內建字體的故事：柊野明朝與游明朝
- Font smoothing, anti-aliasing, and sub-pixel rendering
- Laissez-faire Font Smoothing and Anti-aliasing
- A Closer Look At Font Rendering
Thank you

https://www.chenhuijing.com

@hj_chen

@hj_chen

@huijing