CSS for i18n

By Chen Hui Jing / @hj_chen
Numeronyms

internationalisation

localisation
“Internationalisation is the design and development of a product, application or document that enables easy localisation for target audiences that vary in culture, region, or language.”
What is CSS?

A vocabulary for communicating with the browser.
To inform the browser how we want elements to be styled and laid out.

“CSS is a domain-specific, declarative programming language.”
—Lara Schenck
Why `lang` matters?

- Language-specific behaviour in browsers
- Default font selection
- Search engine optimisation
- Spelling and grammar checks
- Translation
- Non-text readers
- Parsers and scripts

Source: Why use the language attribute?
Basics of using `lang`

Set the default language of the document:

```
<html lang="zh">
```

If you have mixed language content on your page:

```
<p>The fourth animal in the Chinese Zodiac is Rabbit ("span lang="zh">兔子</span").</p>
```
The `:lang()` pseudo-class

A selector that represents an element based on its language

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

A selector that matches elements based on the presence or value of a given selector

```html
[lang="zh"]
/* will match only zh */
```

```html
[lang~="zh"]
/* will match zh, zh-HK, zh-Hans, zhong, zh123...
* basically anything with zh as the first 2 characters */
```

```html
[lang!="zh"]
/* will match zh, zh-HK, zh-Hans, zh-amazing, zh-123 */
```
Some things to note...

The `:lang()` pseudo-class does take care of partial tag matching.

In terms of selector matching, the `=` operator performs a comparison against a given attribute on the element, while the `:lang()` pseudo-class uses the UAs knowledge of the document's semantics for the comparison.

You can always use normal classes or ids, but you would no longer be making use of the convenience of what already exists.
The faux fonts problem

Spotting faux bold

Spotting faux italic

Image source: Cai
Controls if the browser should synthesize a missing typeface or not

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

```
.faux em:lang(zh) {
    font-synthesis: none;
}
```
Glyph differences
**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」。
font-language-override

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

<!--- Macedonian lang code -->

<body lang="mk">
  <h4>Член 9</h4>
  <p>Никој човек нема да биде подложен на произволно апсење, притвор или прогонување.</p>
</body>

body {
  /* Serbian OpenType language tag */
  font-language-override: "SRB";
}

Example lifted from CSS Fonts Module Level 4
If I want [flowers], I’m going to send them to myself.

SÜSSE SOSSEN-KLÖSSE GENIESSEN MASSGEBLICH GEFÄRÄSSIGE PREUSSISCHE NUTZNIESSER.

ΟΥΔ’ΕΝ ΚΑΚ’ΟΝ ΑΜΙΓ’ΕΣ ΚΑΛΟ’Υ.

アイウエオカクケ
Line breaks for inline boxes

If an element \textit{generates zero boxes}, was it \textbf{really there} at all?

\begin{verbatim}
<p class="line-container">If an element <em>generates zero boxes</em>, was it <strong>really there</strong> at all?</p>
\end{verbatim}
## 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
Hyphenation

"If you don't give a lang attribute, you don't get automatic hyphenation."

—Florian Rivoal

Browsers use language-specific dictionaries to figure out where the hyphenation points should be.
すべての人間は、生まれながらにして自由であり、かつ、尊重と権利について平等である。人間は、理性と良心とを授けられており、互いに同胞の精神をもって行動しなければならない。

writing-mode: vertical-lr

writing-mode: vertical-rl
國家籃球協會（英語：National Basketball Association，縮寫：NBA）是北美的男子職業籃球聯盟。

語：Basketball Association。
Right-to-left languages

CSS should not be used for bi-directional styling

“Because directionality is an integral part of the document structure, markup should be used to set the directionality for a document or chunk of information, or to identify places in the text where the Unicode bidirectional algorithm alone is insufficient to achieve desired directionality.”
Not just for East Asian text
Confusing physical directions

Physical directions for writing modes other than horizontal top-to-bottom may be confusing
## Logical properties

<table>
<thead>
<tr>
<th>writing-mode / direction</th>
<th>horizontal-tb</th>
<th>vertical-rl</th>
<th>vertical-lr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ltr</td>
<td>rtl</td>
<td>ltr</td>
</tr>
<tr>
<td><strong>top</strong></td>
<td>inset-block-start</td>
<td>inset-block-start</td>
<td>inset-inline-start</td>
</tr>
<tr>
<td><strong>right</strong></td>
<td>inset-inline-end</td>
<td>inset-inline-start</td>
<td>inset-block-start</td>
</tr>
<tr>
<td><strong>bottom</strong></td>
<td>inset-block-end</td>
<td>inset-block-end</td>
<td>inset-inline-end</td>
</tr>
<tr>
<td><strong>left</strong></td>
<td>inset-inline-start</td>
<td>inset-inline-end</td>
<td>inset-block-end</td>
</tr>
</tbody>
</table>
A comparison of physical and logical directions for borders

The requirements is to have a box with text inside it with the following characteristics:
1. The border color at the top edge of the text must be orange.
2. The border color at the right edge of the text must be grey.
3. The border color at the bottom edge of the text must be blue.
4. The border color at the left edge of the text should be yellow.

Using physical directions requires a modification every time the writing direction changes, whereas using logical properties allows the same properties and values for all six cases.

Physical directions

Logical directions
Lists and counters

Allows us to display ordered lists with international numbering systems

甲、乙、丙、丁、戊、己、庚、辛、壬、癸、甲、甲、甲、乙、丙、
CSS Fizzbuzz

1.
2.
Fizz

4.
Buzz

Fizz

7.
8.
Fizz
Buzz

11.

```css
.fizzbuzz ol li:nth-of-type(3n+3)::marker {
  content: "Fizz"
}
.fizzbuzz ol li:nth-of-type(5n+5)::marker {
  content: "Buzz"
}
.fizzbuzz ol li:nth-of-type(3n+3):nth-of-type(5n+5)::marker {
  content: "FizzBuzz"
}
```
Further reading and references

- CSS for internationalisation
- CSS is a Declarative, Domain-Specific Programming Language
- Why use the language attribute?
- Using the HTML lang attribute
- On Use of the Lang Attribute
- Declaring language in HTML
- Robust Vertical Text Layout
- Unicode Bidirectional Algorithm
Thank you!

https://www.chenhuijing.com

@hj_chen

@hj_chen

@huijing

Font is Crimson Pro by Jacques Le Bailly.