HTML overview

Web development I: Front-end engineering

Terminology

GML → SGML → XML → HTML

Document type declaration

Markup vs Content

Tag vs Element

Attributes

Encoding

Character entities (escaping)

HTML vs XML

UNIVERSITÉ DU LUXEMBOURG

XML: eXtensible Markup Language

Focused on describing *data*

Dynamic: Transport purposes

Any tag is possible

Case-sensitive syntax

Must be well formed

Type safety and verbosity

HTML: HyperText Markup Language

Focused on describing *documents*

Static: Display purposes

Predefined set of tags

Case-insensitive syntax

May have errors

What browsers ignore

Multiple white spaces

Line breaks

Tabs

Unrecognized markup

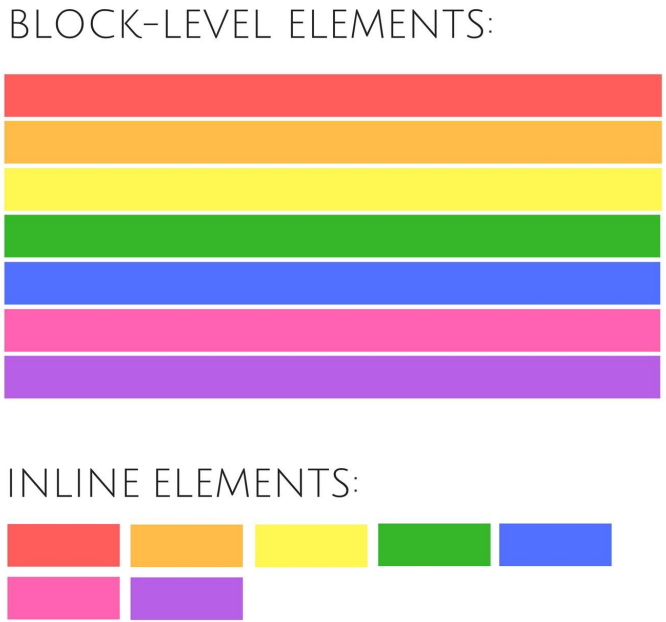
Anything inside comments

Browser defaults

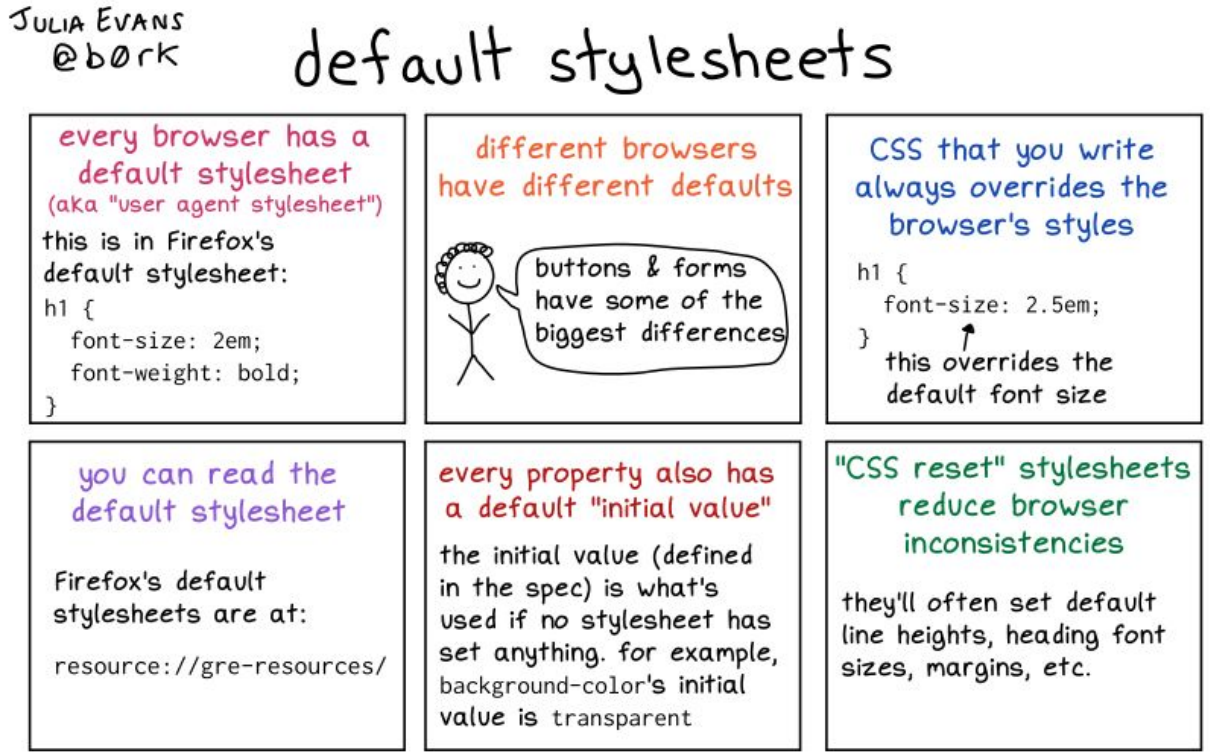
No doctype implies “quirks mode”

Block vs inline elements

- Rendering implications: `width` and `margin`
- Any non-standard tags are considered inline



Browser defaults



Semantic markup

UNIVERSITÉ DU LUXEMBOURG

Provide **meaning** to documents

- Examples: `section`, `main`, `header`, `footer`, `aside`, `nav`

Microdata: https://www.w3.org/TR/microdata/

Microformats: https://microformats.io/

Empty elements

Were introduced to HTML by mistake: presentational markup crept into the language

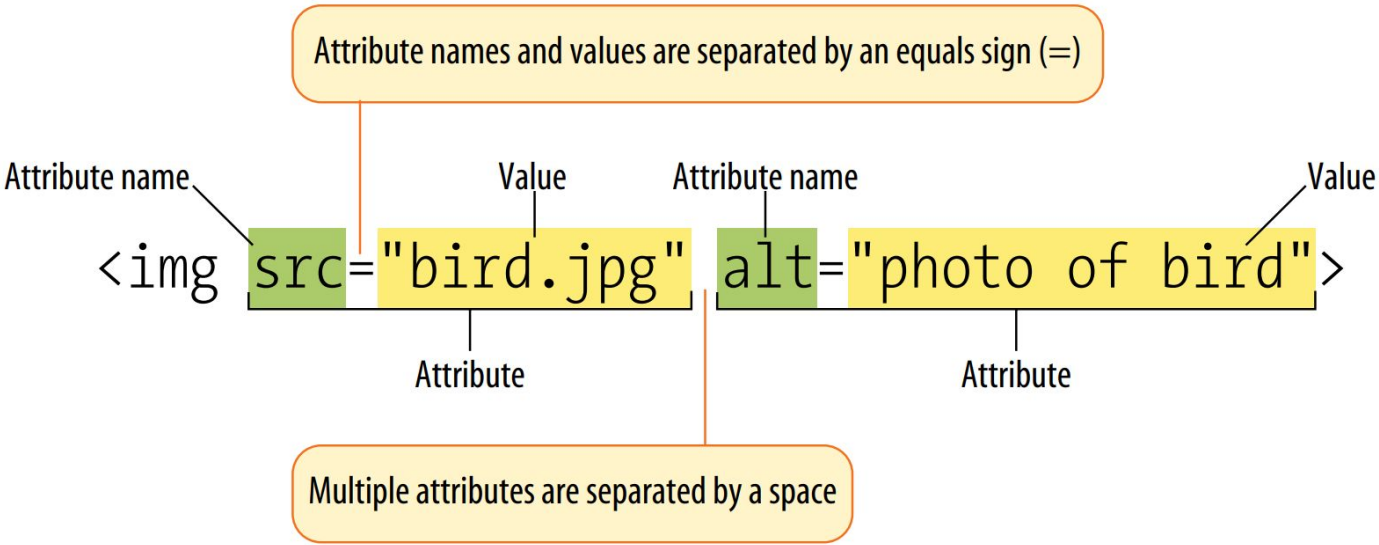
May or may not have attributes

Examples: image, line break, rule, meta tags

Good practice: close them explicitly (e.g. `<br />` instead of `<br>`)

https://www.w3.org/TR/xhtml1/#C_2

Attributes


Attribute names and values are separated by an equals sign (=)
Attribute name
Value
Attribute name
Value
<img src="bird.jpg" alt="photo of bird">
Attribute
Attribute
Multiple attributes are separated by a space

Attributes

Order does not matter but repetition does

Some can be empty: `<input type="text" disabled />`

Some can omit quotes: `<input value=yes />`

Custom attributes are prefixed with **`data-`**: `<input data-hello="101" />`

Special semantics: `id` (unique identifier) and `class` (classifier)

Common **doctypes** (and DTDs)

UNIVERSITÉ DU LUXEMBOURG

HTML 4.01 (strict)

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
   "http://www.w3.org/TR/html4/strict.dtd">
```

XHTML 1.1

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC
  "-//W3C//DTD XHTML 1.1//EN"
  "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
```

HTML5

```
<!DOCTYPE html>
```

Character entities

Some characters are reserved, e.g. <, >, &, "

Escaping sequence is & + word + ; or &# + number + ; or &#x + hex + ;

`<` is displayed as <

`&#169;` is displayed as ©

`&#x000A2;` is displayed as ¢

See https://dev.w3.org/html5/html-author/charref

How many HTML tags are there?

UNIVERSITÉ DU LUXEMBOURG

“144 distinct tag names, out of which 28 are deprecated.”

— https://developer.mozilla.org/en-US/docs/Web/HTML/Element