



香港中文大學
The Chinese University of Hong Kong

HTML

CSCI2720 2022-23 Term 2
Building Web Applications

Dr. Chuck-jeे Chau
chuckjee@cse.cuhk.edu.hk

OUTLINE

- HTML Basics
- Marking up elements
- Hyperlinks
- Encoding special characters

HTML BASICS

- HTML – Hypertext Markup Language
 - It is *not a programming language* but a rendering guideline for software
- The most *fundamental* code web browsers read to generate web contents

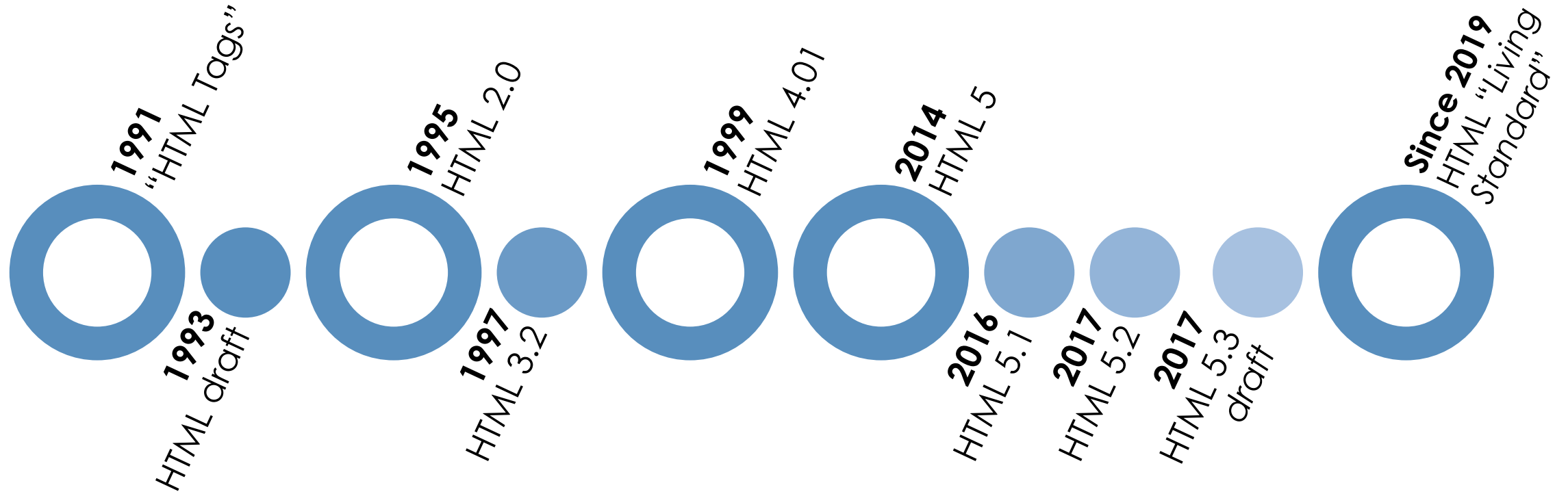


```
ass="col-xs-12 hidden-xs col-sm-7 col-md-7">
id="search_lang_nav">
v id="cuhk-social-channels" class="cuhk-social-cha
os://www.facebook.com/CUHKofficial " target="_blan
lish/images/social-channels/purple-icon-fb.png" al
fficial Facebook Page" /></a> <a
os://www.youtube.com/user/CUHKchannel" target="_bl
s/purple-icon-youtube.pn
nel" /></a> <a
chineseuniversityofhongk
images/social-channels/p
Instagram" /></a> <a
ol/13045 " target="_blan
s/purple-icon-in.png" al
a
ial" target="_blank"><im
s/purple-icon-twitter.pn
r" /></a> <div
ascript:;" class="wechat
annels/purple-icon-wecha
" /></a><div

cons/wechat_CUHKofficial
channel" /><br/>WeChat
ttps://t.me/s/cuhkoffici
images/social-channels/p
am" /></a> <a
icial" target="_blank"><
s/purple-icon-sina-weibo

rm"
hksearch/search.aspx" me
v form="form" class="search-box form">
<input type="hidden" name="domain" value="cuhk.ed
<input type="hidden" name="inurl" value="" />
<input type="hidden" name="enc" value="utf-8" />
<input type="search" placeholder="Search" name="q
searchbox" title="Enter keyword to search CUHK webs
<input type="image" src="/english/images/btn_sear
searchbox_submit" alt="Submit" />
/form>
```

A BRIEF HISTORY OF HTML...



SOME IDEAS ON WEB ORGANIZATIONS...

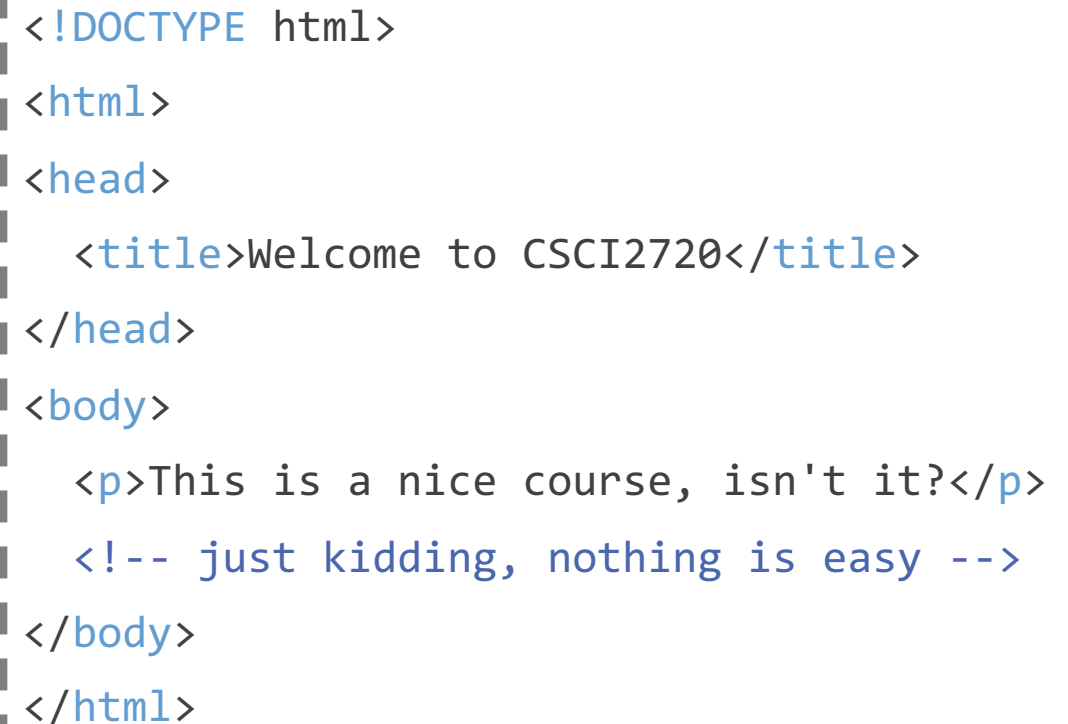
- W3C (*since 1994*)
 - **World Wide Web Consortium**
 - Founded by Tim-Berners Lee
 - the creator of WWW
 - Maintaining standards for WWW
 - Working draft, candidate/proposed recommendations
 - W3C recommendations
- WHATWG (*since 2004*)
 - **Web Hypertext Application Technology Working Group**
 - Founded by people in leading web browser vendors
 - Read:
 - <https://www.w3.org/blog/2019/05/w3c-and-whatwg-to-work-together-to-advance-the-open-web-platform/>

WHY HTML?

- HTML helps you to
 - ***dedicate the roles*** of text or media on the page
 - set up ***hyperlinks*** to allow navigation between pages
- HTML is well supported by web browsers on multiple device platforms, allowing a ***unified*** experience
- Although people rarely write HTML directly, you need to learn basic concepts to generate a page using ***scripts***!

A SIMPLE HTML DOCUMENT

- The **!DOCTYPE** declares the document type
 - "html" represents an HTML5 file
- The **<head>** section contains useful data but not for displaying, such as scripts and stylesheets
- The **<body>** section contains everything to be shown in the browser screen
 - **<!--** and **-->** denotes comments which will be ignored when rendering
- Usually this is saved as a **.html** file

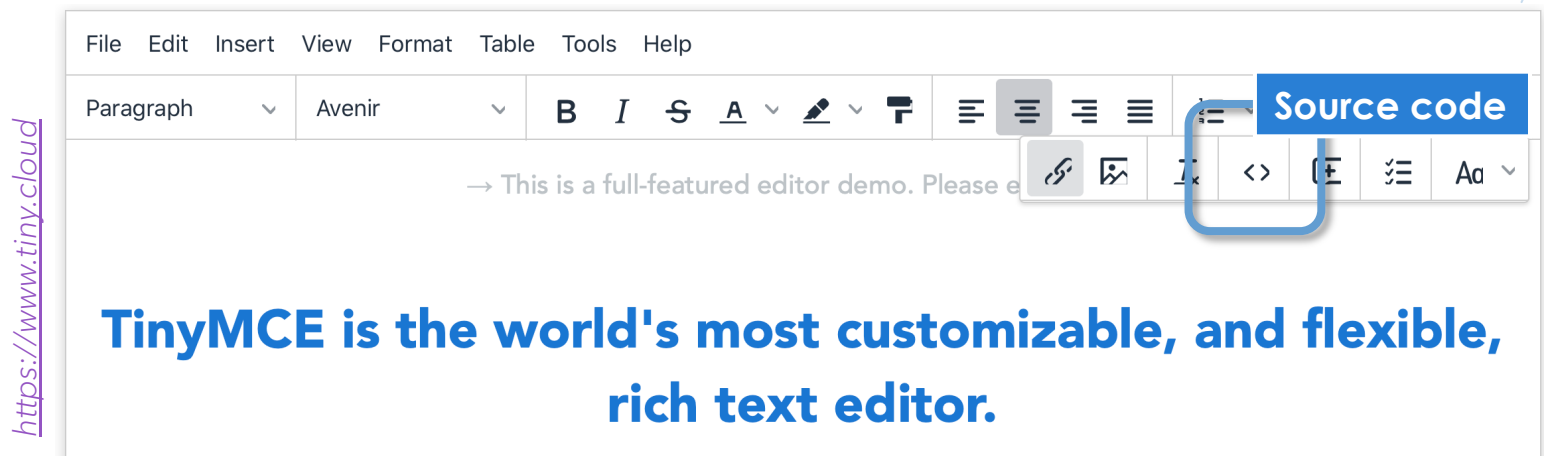


```
<!DOCTYPE html>
<html>
  <head>
    <title>Welcome to CSCI2720</title>
  </head>
  <body>
    <p>This is a nice course, isn't it?</p>
    <!-- just kidding, nothing is easy -->
  </body>
</html>
```

<https://codepen.io/chuckjee/pen/abmjgVJ>

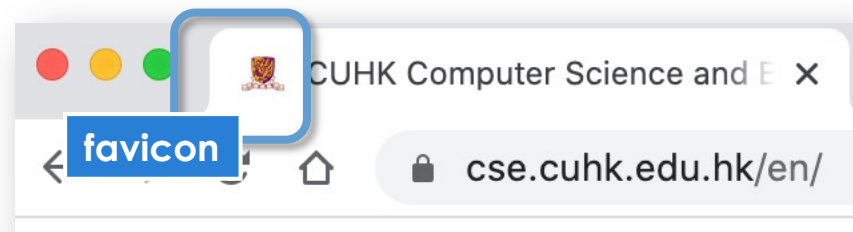
WHERE DO YOU SEE HTML CODE?

- Looking at the source code on any web page
 - Chrome: right click and choose "**View Page Source**"
 - Safari: right click and choose "**Show Page Source**"
- "**Source code**" in some WYSIWYG editors on web



THE HTML HEAD

- Some items are relevant to a web page, but are not contents to be shown in the page
 - Page title and “***favicon***” of a page
 - Stylesheets, scripts or other external files
 - Metadata like keywords for search engines to ***understand*** the page in their way

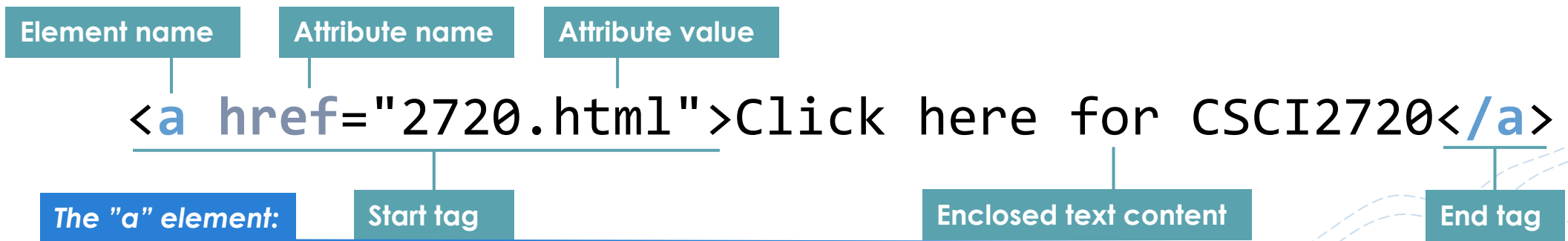


THE HTML BODY

- All contents in the body will be shown in the page
 - *Paragraphs, headings, images, tables, ...*
- You can also create a nice structure to present the contents in a semantic manner, using a header, sections and a footer
- Sometimes executable scripts are put at the end of the HTML body

THE SYNTAX OF HTML ELEMENTS

- All HTML elements building blocks of the web page
 - Whether they are shown or not
- Elements are created using tags in the code
 - Tags may or may not have an attribute
- Note: HTML is **not case-sensitive**, yet recommended for small letter tags



MARKING UP ELEMENTS

- Headings `<h1>`, `<h2>`, ..., `<h6>`
- Paragraph `<p>` and line break `
`
- Formatting
 - Bold ``, italic `<i>`, underline `<u>`
 - Subscript `<sub>`, superscript `<sup>`
 - Pre-formatted `<pre>`

Heading 1

Heading 2

A paragraph with **bold text**, *italic text*, and underlined text
with line break followed by _{subscript} and ^{super}script

```
Here are      some  
              preformatted  
text.
```

MARKING UP ELEMENTS

- Lists

- Ordered list ``
- Unordered list ``
- List items ``

```
<ul>  
<li>Item 1</li>  
<li>Item 2</li>  
</ul>
```

```
<ul type="square">  
<li>Item 1</li>  
<li>Item 2</li>  
</ul>
```

```
<ul type="circle">  
<li>Item 1</li>  
<li>Item 2</li>  
</ul>
```

- Item 1
- Item 2

- Item 1
- Item 2

- Item 1
- Item 2

```
<ol>  
<li>Item 1</li>  
<li>Item 2</li>  
</ol>
```

```
<ol type="A"  
    start="5">  
<li>Item 1</li>  
<li>Item 2</li>  
</ol>
```

```
<ol type="i"  
    start="10">  
<li>Item 1</li>  
<li>Item 2</li>  
</ol>
```

1. Item 1
2. Item 2

E. Item 1
F. Item 2

x. Item 1
xi. Item 2

MARKING UP ELEMENTS

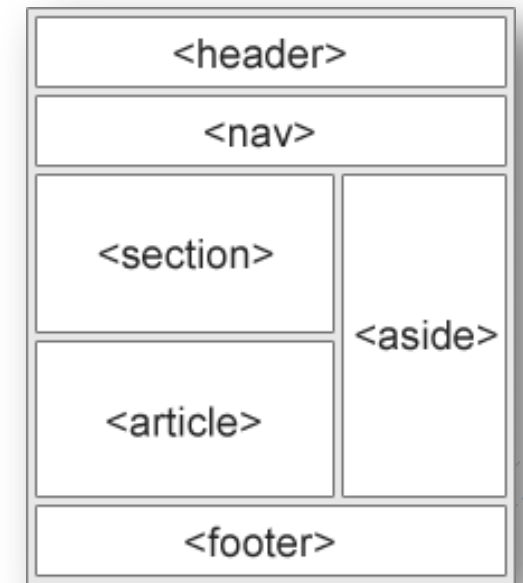
- Tables `<table>`
 - A table is broken into rows `<tr>`
 - A row is broken into data cells `<td>`
 - Optional table header `<th>`

```
<table border="1">  
  <tr>  
    <td>row 1, cell 1</td>  
    <td>row 1, cell 2</td>  
  </tr>  
  <tr>  
    <td>row 2, cell 1</td>  
    <td>row 2, cell 2</td>  
  </tr>  
</table>
```

row 1, cell 1	row 1, cell 2
row 2, cell 1	row 2, cell 2

SEMANTIC ELEMENTS

- In HTML5, it is recommended that the page contents are declared clearly into logical sections
 - e.g., sections **<section>**, navigation bar **<nav>**
- Browsers generally do not define how to render them
 - Easier for search engines and bots to know how data is organized on the page
 - Good anchor points for styling up the page with CSS
 - *Except bold ****, italic ****, etc. which have predefined behaviours in a browser*
- See: https://www.w3schools.com/html/html5_semantic_elements.asp



HYPERLINKS

- The hyperlink allows a “non-linear” manner of hypertext and hypermedia consumption
 - **Inline links**: pointing to another file in the same server, or files on another web server
 - **anchors**: pointing to another part/section in the same file
- Usually displayed in different colours than normal text, depending on whether the link has been visited or not

HYPERLINKS

- The `<a>` element
 - `href` attribute → what to point to
 - `target` attribute → where to open, e.g. `"_blank"` opens the link in a new tab/window
 - e.g., `Webapp` will open the files `csci2720.html` in a new tab/window
- Defining a fragment name using an id could be useful, e.g.
 - We have `<h2 id="intro">Introduction</h2>` in `csci2720.html`
 - The link can point directly to it as:
`Introduction`

ABSOLUTE VS. RELATIVE PATHS

Absolute paths, e.g.

`http://www.cuhk.edu.hk/english/index.html`

- Using a complete URL (uniform resource locator)
 - Protocol (`http`)
 - Domain (`www.cuhk.edu.hk`)
 - Port (`80`, not typed by default)
 - Path (`/english/`)
 - Filename (`index.html`)

Relative paths, e.g.

`../hello.html`

- Using the current document as reference
- e.g. We are at the address `http://www.cuhk.edu.hk/english/index.html`
- `` brings us to the "chinese" directory under "../" parent directory
- What about `href="/index2.html"`?

INCLUDING IMAGES

- Modern browsers support generally lots of image types, usually using ``
 - e.g., ``
- See: https://developer.mozilla.org/en-US/docs/Web/Media/Formats/Image_types
- The special tag `<svg>` can be used for the Scalar Vector Graphics
 - Specifying contents of a graphic using elements, e.g., `<circle>`, `<line>`
- People also use `<picture>` for detailed control on responsiveness

EMBEDDING AUDIO AND VIDEO

- The relatively newer elements of `<audio>` and `<video>` adds native multimedia support into browsers
- Since there are too many multimedia file formats out there, you can use multiple `<source>` tags to point to multiple files, e.g.,

```
<video control width="500">  
  <source src="2720ver1.mp4" type="mp4">  
  <source src="2720ver2.webm" type="webm">  
  <p>Your browser isn't supported!</p>  
</video> <!-- lines observed one by one in order -->
```

ENCODING SPECIAL CHARACTERS

- Browsers doesn't like to see `<` or `>` in the text as they look too much like HTML tags, e.g.,

```
<p>Hello, I believe x < y and y > z.</p>
```

 - Modern browsers may be able to guess correctly, but who want to risk losing some customers seeing your page?
- `<` should be typed as `<` and `>` should be as `>` in the HTML file
 - These are called "HTML entities" and there are a list of them
 - See: <https://dev.w3.org/html5/html-author/charref>

HANDLING SPACE

- By default, more than one consecutive whitespace (space, new line, tab, etc.) in an HTML file will be regarded as one, e.g.
 - `<p>Hello World</p>` will be rendered as:
→ Hello World
- ` ` is the “non-breaking space” for inserting multiple whitespace, or avoiding line breaks
- Whitespace is preserved in the `<pre>` environment

THERE ARE MORE TO LEARN!

- We have only gone through very superficial features in the HTML language and features
 - These are cornerstones which you will see again and again
- We will learn more throughout the course
- Check out HTML Cheatsheets, e.g.,
 - <https://www.wpkube.com/html5-cheat-sheet/>



HTML Living Standard

<https://html.spec.whatwg.org>

w3schools.com HTML5 Tutorial

<https://www.w3schools.com/html>

MDN HTML Guides and tutorials

<https://developer.mozilla.org/en-US/docs/Learn/HTML>

READ FURTHER...