




HTML

The code for the web
Summary lecture slides



Relevance

- HTML is the base of web pages
- Web pages make up the world wide web
- Knowing some very basic HTML is useful

History

- Created around 1991 by Tim Berners-Lee (at CERN)
- HTML stands for HyperText Markup Language
- Hypertext is (computer) text with “hyperlinks” that allows you to instantly go to a referenced source
- HTML is a “markup language” (uses code to mark up text)
- Browsers “understand” HTML and “parses” it to display pages with text, images and other media
- Turned out to be a really good idea
- *Not* a programming language - you don't program a web page, you write markup for it

Code up-front! (color added as sugar)

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="utf-8">
    <title>This is shown as the title of the window/tab</title>
  </head>
  <body>
    <p>A paragraph of text looks like this.</p>
  </body>
</html>
```

Markup

- HTML consists of elements that enclose some text or other elements
- An element uses an opening tag, like `<p>` some content, and a closing tag, like `</p>` when marking up the content
- Some elements are empty and can be shortend, like `
` (linebreak)
- Elements can be nested like

```
<body>  
  <p>Paragraph text here.</p>  
</body>
```
- Write tags in lowercase

Elements part 1

```
<p>                ← opening tag, element name is “p”  
Hello there.      ← content, in this case text  
</p>             ← closing tag
```

Elements part 2

<code><p></code>	← opening tag
<code>Hello there.</code>	← content, in this case text and a nested element
<code></p></code>	← closing tag

When nesting, you need to close the nested elements within the outer element.

This is wrong: `<p>Some text</p>`

Elements part 3 - block elements

```
<p>A paragraph</p><p>A new paragraph</p>
```

- Block elements appear by themselves surrounded by newlines
- Used for structure - building paragraphs, headlines, lists etc

Elements part 4 - inline elements

`<p>A paragraph with a nested inline element</p>`

- Inline elements are nested inside block elements
- Are not surrounded by newlines, and typically markup a small part of text inside some block element
 - styles like `bold text` or `emphasis like italics`
 - hyperlinks:

`<p>`

This text contains a `hyperlink`

`</p>`

Elements part 5 - Attributes

- Some (most) elements have attributes with extra information
- Attributes have a name and a value and occur within the opening tag:
`Link text`
 - **href** and **title** the attribute names and the values come after an equal sign and should be enclosed by double quotes
- Boolean-valued attributes don't need a value (exception to the rule):
`<video src="apa.mpg" muted></video>`

Full page revisited - identify elements and attr.

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="utf-8">
    <title>This is shown as the title of the window/tab</title>
  </head>
  <body>
    <p>A paragraph of text looks like this.</p>
    <p>A link <a href="index.html">link</a> looks like this.</p>
  </body>
</html>
```

Identify: Elements, tags, attributes and content.

Save the code as a file called test.html and open that file in a browser. Can you identify what you see in relation to the code?

Whitespace

HTML treats whitespace between elements as one single whitespace (newline for block elements, and one space for inline elements).

This:

```
<p>   Hello           whitespace in           HTML   </p>
```

Will look like this:

Hello whitespace in HTML

Special characters

- HTML is a markup language using tags to describe elements
<p>Content</p>
- But how do you produce content that has < or > as part of the content (not intended to be part of tags, but intended to be printed)?
- < is produced by <
- > is produced by >
- " is produced by "
- ' is produced by '
- & is produced by &

There are more "[HTML entities](#)".

Comments in the source code

- Sometimes you want to leave a comment in the source code intended for yourself or other developers - but not intended to be processed by e.g. the browsers
- Everything between `<!--` and `-->` is ignored by browsers (and other HTML processors)

```
<p>This is a paragraph <!-- but this won't show up --> </p>
```

```
<!--
```

```
Nothing here will “count” as part  
of the document
```

```
-->
```

Doctype

The very first thing in an HTML document, is the doctype declaration:

```
<!DOCTYPE html>
```

This simply declares that the document is of type HTML.

Since HTML5, the above is enough.

The head of the document

Remember this?

```
<!DOCTYPE html>  
<html lang="en">  
  <head>  
    <meta charset="utf-8">  
    <title>This is shown as the title of the window/tab</title>  
  </head>  
  <body>
```

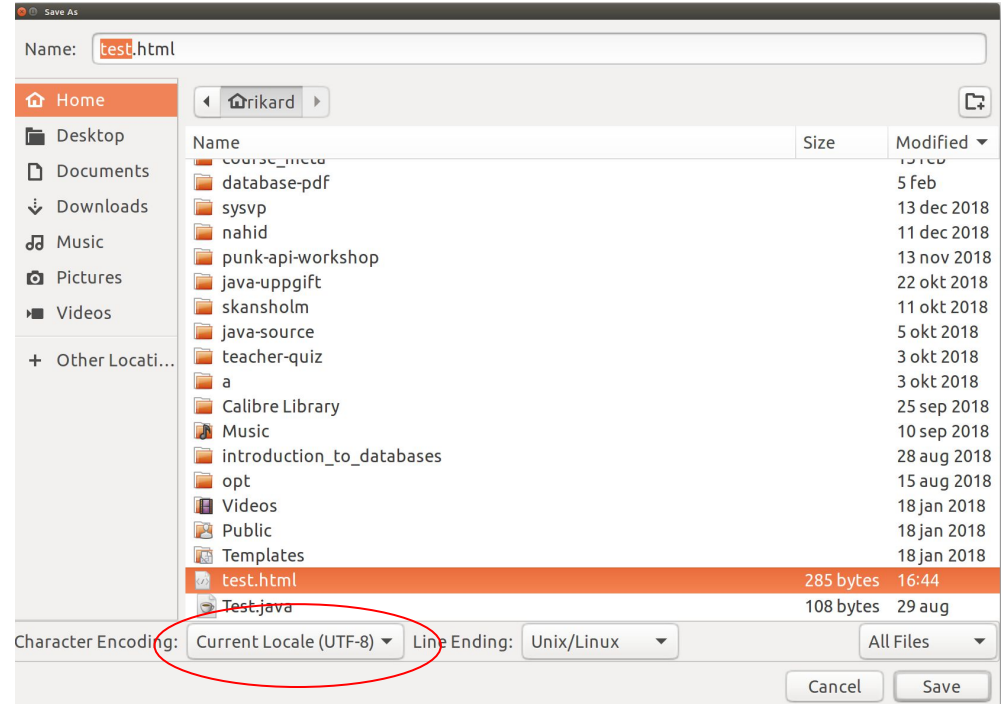
...

You put metadata in the head element

- `<meta charset="utf-8">`
- `<title>This is shown as the title of the window/tab</title>`
- You can put other metadata in the head, and even stylesheet instructions in the CSS (cascading style sheets) language
 - CSS is outside the scope of this course materials
- If you are using e.g. Swedish characters, you need to specify the charset as utf-8 and save the file in this encoding, using the meta tag with the charset attribute as shown above.
- utf-8 is a widely used charset on the internet and can encode the whole unicode character set, which consists of most written languages and also a lot of graphical symbols
- Most editors can save in utf-8 (many will detect Swedish characters and default to utf-8 for you)

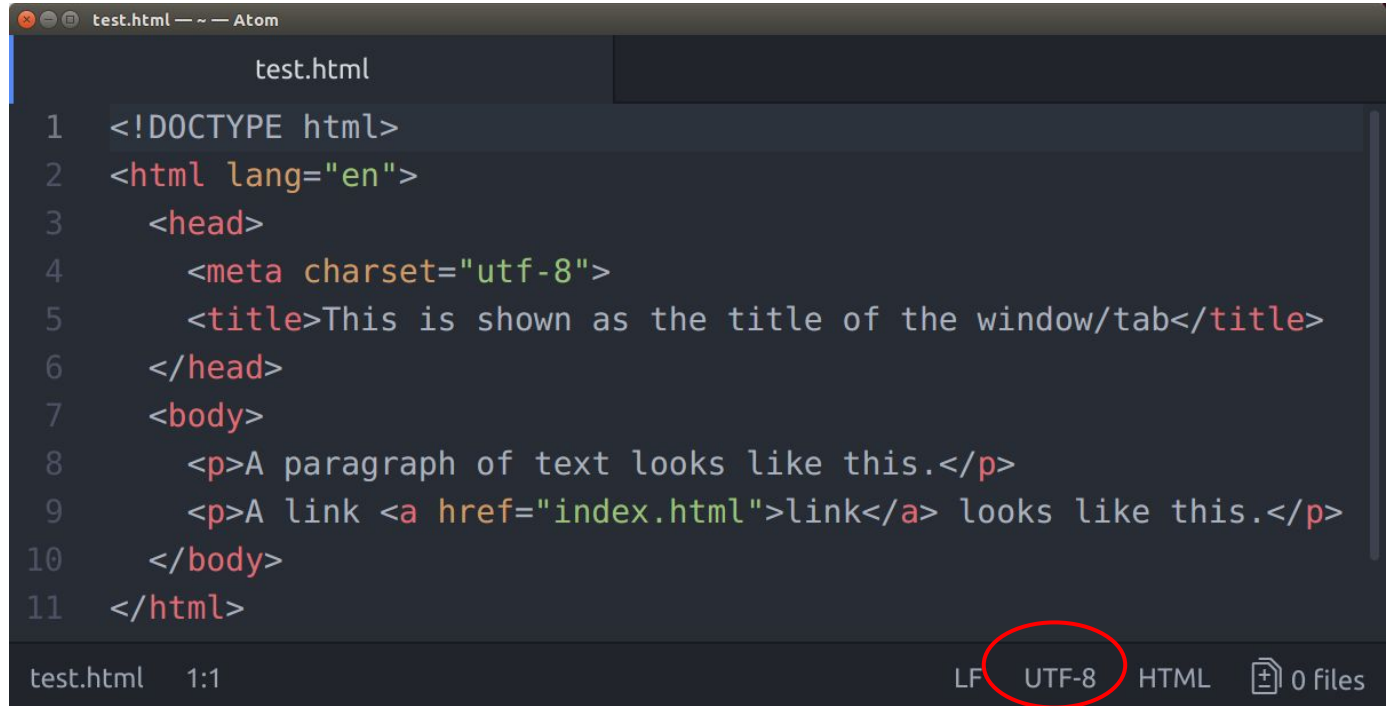
UTF-8 - Gedit

Gedit - save as - UTF-8
(Uses system default)



UTF-8 - Atom

You can click on the button for encoding in Atom, to change encoding...



```
test.html — ~ — Atom
test.html
1 <!DOCTYPE html>
2 <html lang="en">
3   <head>
4     <meta charset="utf-8">
5     <title>This is shown as the title of the window/tab</title>
6   </head>
7   <body>
8     <p>A paragraph of text looks like this.</p>
9     <p>A link <a href="index.html">link</a> looks like this.</p>
10  </body>
11 </html>
```

test.html 1:1 LF UTF-8 HTML 0 files

UTF-8 - Emacs

C-x C-m f

Ctrl-X, then

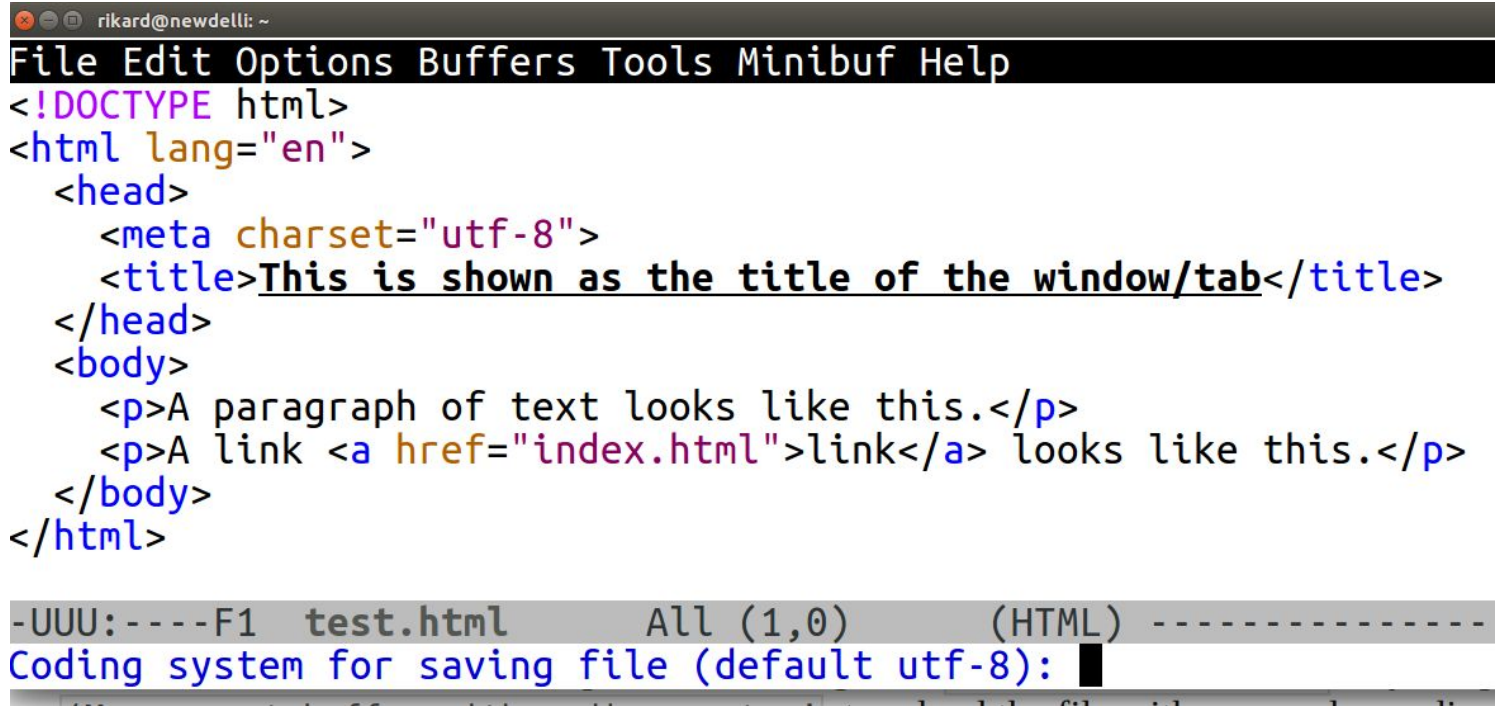
Ctrl-M, then

a single f

lets you

change

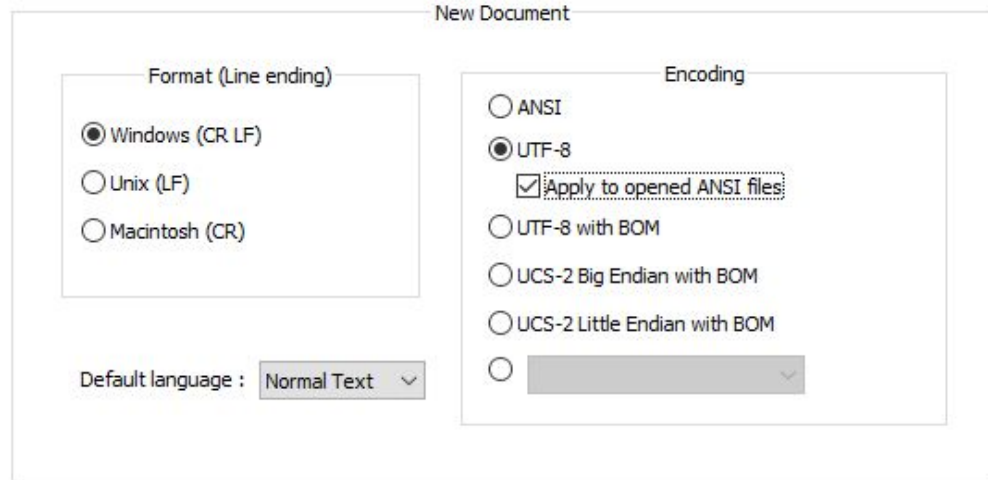
encoding...



```
rikard@newdelli: ~  
File Edit Options Buffers Tools Minibuf Help  
<!DOCTYPE html>  
<html lang="en">  
  <head>  
    <meta charset="utf-8">  
    <title>This is shown as the title of the window/tab</title>  
  </head>  
  <body>  
    <p>A paragraph of text looks like this.</p>  
    <p>A link <a href="index.html">link</a> looks like this.</p>  
  </body>  
</html>  
  
-UUU:----F1 test.html All (1,0) (HTML) -----  
Coding system for saving file (default utf-8): █
```

UTF-8 - Notepad++

Preferences



Close

What elements are there?

- Tons of more elements
- We won't list them all here, but there are plenty of links with resources on the wiki!
- Basic stuff to learn (find out) include:
- paragraphs
- emphasis, strong, linebreak
- headers (level 1, 2, 3 etc)
- links
- lists (making e.g. a bullet list in html)

What about CSS?

- Cascading style sheets (CSS) is a way to style your document, by keeping the style separated from the content
- It's considered bad style to mix structure/content and style
- You can learn about CSS from various online resources
 - https://developer.mozilla.org/en-US/docs/Learn/CSS/First_steps