

# Chapter 6 More about HTML

Tuesday, September 20, 2016 10:27 PM

## HTML Entities

Some characters are reserved in HTML.

If you use the less than (<) or greater than (>) signs in your text, the browser might mix them with tags.

Character entities are used to display reserved characters in HTML.

A character entity looks like this:

`&entity_name;`

OR

`&#entity_number;`

To display a less than sign (<) we must write: **&lt;** or **&#60;**;

From <[http://www.w3schools.com/html/html\\_entities.asp](http://www.w3schools.com/html/html_entities.asp)>

## Non-breaking Space

A common character entity used in HTML is the non-breaking space: **&nbsp;**;

From <[http://www.w3schools.com/html/html\\_entities.asp](http://www.w3schools.com/html/html_entities.asp)>

## Some Other Useful HTML Character Entities

Result	Description	Entity Name	Entity Number
	non-breaking space	&nbsp;	&#160;
<	less than	&lt;	&#60;
>	greater than	&gt;	&#62;
&	ampersand	&amp;	&#38;
"	double quotation mark	&quot;	&#34;
'	single quotation mark (apostrophe)	&apos;	&#39;
¢	cent	&cent;	&#162;
£	pound	&pound;	&#163;
¥	yen	&yen;	&#165;
€	euro	&euro;	&#8364;
©	copyright	&copy;	&#169;
®	registered trademark	&reg;	&#174;

From <[http://www.w3schools.com/html/html\\_entities.asp](http://www.w3schools.com/html/html_entities.asp)>

## HTML Symbol Entities

HTML entities were described in the previous chapter.

Many mathematical, technical, and currency symbols, are not present on a normal keyboard.

To add such symbols to an HTML page, you can use an HTML entity name.

If no entity name exists, you can use an entity number, a decimal, or hexadecimal reference.

## Example

```
<p>I will display &euro;</p>
<p>I will display &#8364;</p>
<p>I will display &#x20AC;</p>
```

From [http://www.w3schools.com/html/html\\_symbols.asp](http://www.w3schools.com/html/html_symbols.asp)

## Some Other Entities Supported by HTML

Char	Number	Entity	Description
©	&#169;	&copy;	COPYRIGHT SIGN
®	&#174;	&reg;	REGISTERED SIGN
€	&#8364;	&euro;	EURO SIGN
™	&#8482;	&trade;	TRADEMARK
←	&#8592;	&larr;	LEFTWARDS ARROW

From [http://www.w3schools.com/html/html\\_symbols.asp](http://www.w3schools.com/html/html_symbols.asp)

## The HTML charset Attribute

To display an HTML page correctly, a web browser must know the character set used in the page. This is specified in the <meta> tag:

### For HTML4:

```
<meta http-equiv="Content-Type" content="text/html; charset=ISO-8859-1">
```

### For HTML5:

```
<meta charset="UTF-8">
```

From [http://www.w3schools.com/html/html\\_charset.asp](http://www.w3schools.com/html/html_charset.asp)

## Common URL Schemes

The table below lists some common schemes:

Scheme	Short for	Used for
http	HyperText Transfer Protocol	Common web pages. Not encrypted
https	Secure HyperText Transfer Protocol	Secure web pages. Encrypted
ftp	File Transfer Protocol	Downloading or uploading files
file		A file on your computer

From [http://www.w3schools.com/html/html\\_urlencode.asp](http://www.w3schools.com/html/html_urlencode.asp)

## URL Encoding

URLs can only be sent over the Internet using the [ASCII character-set](#). If a URL contains characters

outside the ASCII set, the URL has to be converted.

URL encoding converts non-ASCII characters into a format that can be transmitted over the Internet.

URL encoding replaces non-ASCII characters with a "%" followed by hexadecimal digits.

URLs cannot contain spaces. URL encoding normally replaces a space with a plus (+) sign, or %2

[http://www.w3schools.com/html/demo\\_form\\_submit.asp?text=Hello+G%C3%BCnter](http://www.w3schools.com/html/demo_form_submit.asp?text=Hello+G%C3%BCnter)

XHTML is HTML written as XML.

## What Is XHTML?

- XHTML stands for **EX**tensible **HyperText Markup Language**
- XHTML is almost identical to HTML
- XHTML is stricter than HTML
- XHTML is HTML defined as an XML application
- XHTML is supported by all major browsers

## Why XHTML?

Many pages on the internet contain "bad" HTML.

This HTML code works fine in most browsers (even if it does not follow the HTML rules):

```
<html>
<head>
  <title>This is bad HTML</title>

  <body>
    <h1>Bad HTML
    <p>This is a paragraph
  </body>
```

From [http://www.w3schools.com/html/html\\_xhtml.asp](http://www.w3schools.com/html/html_xhtml.asp)

## <!DOCTYPE ....> Is Mandatory

From [http://www.w3schools.com/html/html\\_xhtml.asp](http://www.w3schools.com/html/html_xhtml.asp)

## XHTML Elements Must Be Properly Nested

In HTML, some elements can be improperly nested within each other, like this:

```
<b><i>This text is bold and italic</b></i>
```

From [http://www.w3schools.com/html/html\\_xhtml.asp](http://www.w3schools.com/html/html_xhtml.asp)