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;**

**Advantage of using an entity name:** An entity name is easy to remember.  
**Disadvantage of using an entity name:** Browsers may not support all entity names, but the support for numbers is good.

Non-breaking Space

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

A non-breaking space is a space that will not break into a new line.

Two words separated by a non-breaking space will stick together (not break into a new line). This is handy when breaking the words might be disruptive.

Examples:

* § 10
* 10 km/h
* 10 PM

Another common use of the non-breaking space is to prevent that browsers truncate spaces in HTML pages.

If you write 10 spaces in your text, the browser will remove 9 of them. To add real spaces to your text, you can use the**&nbsp;** character entity.

The non-breaking hyphen ([&#8209;](http://www.w3schools.com/charsets/ref_utf_punctuation.asp)) lets you use a hyphen character (‑) that won't break.

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; |
| ¢ | cent | &cent; | &#162; |
| £ | pound | &pound; | &#163; |
| ¥ | yen | &yen; | &#165; |
| € | euro | &euro; | &#8364; |
| © | copyright | &copy; | &#169; |
| ® | registered trademark | &reg; | &#174; |

**Note:** Entity names are case sensitive.

Combining Diacritical Marks

A diacritical mark is a "glyph" added to a letter.

Some diacritical marks, like grave (  ̀) and acute (  ́) are called accents.

Diacritical marks can appear both above and below a letter, inside a letter, and between two letters.

Diacritical marks can be used in combination with alphanumeric characters, to produce a character that is not present in the character set (encoding) used in the page.

Here are some examples:

|  |  |  |  |
| --- | --- | --- | --- |
| **Mark** | **Character** | **Construct** | **Result** |
| ̀ | a | a&#768; | à |
| ́ | a | a&#769; | á |
| ̂ | a | a&#770; | â |
| ̃ | a | a&#771; | ã |
| ̀ | O | O&#768; | Ò |
| ́ | O | O&#769; | Ó |
| ̂ | O | O&#770; | Ô |
| ̃ | O | O&#771; | Õ |

You will see more HTML symbols in the next chapter of this tutorial.

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.

## Some Mathematical Symbols Supported by HTML

|  |  |  |  |
| --- | --- | --- | --- |
| **Char** | **Number** | **Entity** | **Description** |
| ∀ | &#8704; | &forall; | FOR ALL |
| ∂ | &#8706; | &part; | PARTIAL DIFFERENTIAL |
| ∃ | &#8707; | &exist; | THERE EXISTS |
| ∅ | &#8709; | &empty; | EMPTY SETS |
| ∇ | &#8711; | &nabla; | NABLA |
| ∈ | &#8712; | &isin; | ELEMENT OF |
| ∉ | &#8713; | &notin; | NOT AN ELEMENT OF |
| ∋ | &#8715; | &ni; | CONTAINS AS MEMBER |
| ∏ | &#8719; | &prod; | N-ARY PRODUCT |
| ∑ | &#8721; | &sum; | N-ARY SUMMATION |

[Full Math Reference](http://www.w3schools.com/charsets/ref_utf_math.asp)

## Some Greek Letters Supported by HTML

|  |  |  |  |
| --- | --- | --- | --- |
| **Char** | **Number** | **Entity** | **Description** |
| Α | &#913; | &Alpha; | GREEK CAPITAL LETTER ALPHA |
| Β | &#914; | &Beta; | GREEK CAPITAL LETTER BETA |
| Γ | &#915; | &Gamma; | GREEK CAPITAL LETTER GAMMA |
| Δ | &#916; | &Delta; | GREEK CAPITAL LETTER DELTA |
| Ε | &#917; | &Epsilon; | GREEK CAPITAL LETTER EPSILON |
| Ζ | &#918; | &Zeta; | GREEK CAPITAL LETTER ZETA |

[Full Greek Reference](http://www.w3schools.com/charsets/ref_utf_greek.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 |
| ↑ | &#8593; | &uarr; | UPWARDS ARROW |
| → | &#8594; | &rarr; | RIGHTWARDS ARROW |
| ↓ | &#8595; | &darr; | DOWNWARDS ARROW |
| ♠ | &#9824; | &spades; | BLACK SPADE SUIT |
| ♣ | &#9827; | &clubs; | BLACK CLUB SUIT |
| ♥ | &#9829; | &hearts; | BLACK HEART SUIT |
| ♦ | &#9830; | &diams; | BLACK DIAMOND SUIT |

## What is Character Encoding?

ASCII was the first **character encoding standard** (also called character set). ASCII defined 127 different alphanumeric characters that could be used on the internet: numbers (0-9), English letters (A-Z), and some special characters like ! $ + - ( ) @ < > .

ANSI (Windows-1252) was the original Windows character set, with support for 256 different character codes.

ISO-8859-1 was the default character set for HTML 4. This character set also supported 256 different character codes.

Because ANSI and ISO-8859-1 were so limited, the default character encoding was changed to UTF-8 in HTML5.

UTF-8 (Unicode) covers almost all of the characters and symbols in the world.

All HTML 4 processors also support UTF-8 encoding.

## 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">

If a browser detects ISO-8859-1 in a web page, it defaults to ANSI, because ANSI is identical to ISO-8859-1 except that ANSI has 32 extra characters.

## Differences Between Character Sets

The following table displays the differences between the character sets described above:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Numb** | **ASCII** | **ANSI** | **8859** | **UTF-8** | **Description** |
| 32 |  |  |  |  | space |
| 33 | ! | ! | ! | ! | exclamation mark |
| 34 | " | " | " | " | quotation mark |
| 35 | # | # | # | # | number sign |
| 36 | $ | $ | $ | $ | dollar sign |
| 37 | % | % | % | % | percent sign |
| 38 | & | & | & | & | ampersand |
| 39 | ' | ' | ' | ' | apostrophe |
| 40 | ( | ( | ( | ( | left parenthesis |
| 41 | ) | ) | ) | ) | right parenthesis |
| 42 | \* | \* | \* | \* | asterisk |
| 43 | + | + | + | + | plus sign |
| 44 | , | , | , | , | comma |
| 45 | - | - | - | - | hyphen-minus |
| 46 | . | . | . | . | full stop |
| 47 | / | / | / | / | solidus |
| 48 | 0 | 0 | 0 | 0 | digit zero |
| 49 | 1 | 1 | 1 | 1 | digit one |
| 50 | 2 | 2 | 2 | 2 | digit two |
| 51 | 3 | 3 | 3 | 3 | digit three |
| 52 | 4 | 4 | 4 | 4 | digit four |
| 53 | 5 | 5 | 5 | 5 | digit five |
| 54 | 6 | 6 | 6 | 6 | digit six |
| 55 | 7 | 7 | 7 | 7 | digit seven |
| 56 | 8 | 8 | 8 | 8 | digit eight |
| 57 | 9 | 9 | 9 | 9 | digit nine |
| 58 | : | : | : | : | colon |
| 59 | ; | ; | ; | ; | semicolon |
| 60 | < | < | < | < | less-than sign |
| 61 | = | = | = | = | equals sign |
| 62 | > | > | > | > | greater-than sign |
| 63 | ? | ? | ? | ? | question mark |
| 64 | @ | @ | @ | @ | commercial at |
| 65 | A | A | A | A | Latin capital letter A |
| 66 | B | B | B | B | Latin capital letter B |
| 67 | C | C | C | C | Latin capital letter C |
| 68 | D | D | D | D | Latin capital letter D |
| 69 | E | E | E | E | Latin capital letter E |
| 70 | F | F | F | F | Latin capital letter F |
| 71 | G | G | G | G | Latin capital letter G |
| 72 | H | H | H | H | Latin capital letter H |
| 73 | I | I | I | I | Latin capital letter I |
| 74 | J | J | J | J | Latin capital letter J |
| 75 | K | K | K | K | Latin capital letter K |
| 76 | L | L | L | L | Latin capital letter L |
| 77 | M | M | M | M | Latin capital letter M |
| 78 | N | N | N | N | Latin capital letter N |
| 79 | O | O | O | O | Latin capital letter O |
| 80 | P | P | P | P | Latin capital letter P |
| 81 | Q | Q | Q | Q | Latin capital letter Q |
| 82 | R | R | R | R | Latin capital letter R |
| 83 | S | S | S | S | Latin capital letter S |
| 84 | T | T | T | T | Latin capital letter T |
| 85 | U | U | U | U | Latin capital letter U |
| 86 | V | V | V | V | Latin capital letter V |
| 87 | W | W | W | W | Latin capital letter W |
| 88 | X | X | X | X | Latin capital letter X |
| 89 | Y | Y | Y | Y | Latin capital letter Y |
| 90 | Z | Z | Z | Z | Latin capital letter Z |
| 91 | [ | [ | [ | [ | left square bracket |
| 92 | \ | \ | \ | \ | reverse solidus |
| 93 | ] | ] | ] | ] | right square bracket |
| 94 | ^ | ^ | ^ | ^ | circumflex accent |
| 95 | \_ | \_ | \_ | \_ | low line |
| 96 | ` | ` | ` | ` | grave accent |
| 97 | a | a | a | a | Latin small letter a |
| 98 | b | b | b | b | Latin small letter b |
| 99 | c | c | c | c | Latin small letter c |
| 100 | d | d | d | d | Latin small letter d |
| 101 | e | e | e | e | Latin small letter e |
| 102 | f | f | f | f | Latin small letter f |
| 103 | g | g | g | g | Latin small letter g |
| 104 | h | h | h | h | Latin small letter h |
| 105 | i | i | i | i | Latin small letter i |
| 106 | j | j | j | j | Latin small letter j |
| 107 | k | k | k | k | Latin small letter k |
| 108 | l | l | l | l | Latin small letter l |
| 109 | m | m | m | m | Latin small letter m |
| 110 | n | n | n | n | Latin small letter n |
| 111 | o | o | o | o | Latin small letter o |
| 112 | p | p | p | p | Latin small letter p |
| 113 | q | q | q | q | Latin small letter q |
| 114 | r | r | r | r | Latin small letter r |
| 115 | s | s | s | s | Latin small letter s |
| 116 | t | t | t | t | Latin small letter t |
| 117 | u | u | u | u | Latin small letter u |
| 118 | v | v | v | v | Latin small letter v |
| 119 | w | w | w | w | Latin small letter w |
| 120 | x | x | x | x | Latin small letter x |
| 121 | y | y | y | y | Latin small letter y |
| 122 | z | z | z | z | Latin small letter z |
| 123 | { | { | { | { | left curly bracket |
| 124 | | | | | | | | | vertical line |
| 125 | } | } | } | } | right curly bracket |
| 126 | ~ | ~ | ~ | ~ | tilde |
| 127 | DEL |  |  |  |  |
| 128 |  | € |  |  | euro sign |
| 129 |  |  |  |  | NOT USED |
| 130 |  | ‚ |  |  | single low-9 quotation mark |
| 131 |  | ƒ |  |  | Latin small letter f with hook |
| 132 |  | „ |  |  | double low-9 quotation mark |
| 133 |  | … |  |  | horizontal ellipsis |
| 134 |  | † |  |  | dagger |
| 135 |  | ‡ |  |  | double dagger |
| 136 |  | ˆ |  |  | modifier letter circumflex accent |
| 137 |  | ‰ |  |  | per mille sign |
| 138 |  | Š |  |  | Latin capital letter S with caron |
| 139 |  | ‹ |  |  | single left-pointing angle quotation mark |
| 140 |  | Œ |  |  | Latin capital ligature OE |
| 141 |  |  |  |  | NOT USED |
| 142 |  | Ž |  |  | Latin capital letter Z with caron |
| 143 |  |  |  |  | NOT USED |
| 144 |  |  |  |  | NOT USED |
| 145 |  | ‘ |  |  | left single quotation mark |
| 146 |  | ’ |  |  | right single quotation mark |
| 147 |  | “ |  |  | left double quotation mark |
| 148 |  | ” |  |  | right double quotation mark |
| 149 |  | • |  |  | bullet |
| 150 |  | – |  |  | en dash |
| 151 |  | — |  |  | em dash |
| 152 |  | ˜ |  |  | small tilde |
| 153 |  | ™ |  |  | trade mark sign |
| 154 |  | š |  |  | Latin small letter s with caron |
| 155 |  | › |  |  | single right-pointing angle quotation mark |
| 156 |  | œ |  |  | Latin small ligature oe |
| 157 |  |  |  |  | NOT USED |
| 158 |  | ž |  |  | Latin small letter z with caron |
| 159 |  | Ÿ |  |  | Latin capital letter Y with diaeresis |
| 160 |  |  |  |  | no-break space |
| 161 |  | ¡ | ¡ | ¡ | inverted exclamation mark |
| 162 |  | ¢ | ¢ | ¢ | cent sign |
| 163 |  | £ | £ | £ | pound sign |
| 164 |  | ¤ | ¤ | ¤ | currency sign |
| 165 |  | ¥ | ¥ | ¥ | yen sign |
| 166 |  | ¦ | ¦ | ¦ | broken bar |
| 167 |  | § | § | § | section sign |
| 168 |  | ¨ | ¨ | ¨ | diaeresis |
| 169 |  | © | © | © | copyright sign |
| 170 |  | ª | ª | ª | feminine ordinal indicator |
| 171 |  | « | « | « | left-pointing double angle quotation mark |
| 172 |  | ¬ | ¬ | ¬ | not sign |
| 173 |  |  |  |  | soft hyphen |
| 174 |  | ® | ® | ® | registered sign |
| 175 |  | ¯ | ¯ | ¯ | macron |
| 176 |  | ° | ° | ° | degree sign |
| 177 |  | ± | ± | ± | plus-minus sign |
| 178 |  | ² | ² | ² | superscript two |
| 179 |  | ³ | ³ | ³ | superscript three |
| 180 |  | ´ | ´ | ´ | acute accent |
| 181 |  | µ | µ | µ | micro sign |
| 182 |  | ¶ | ¶ | ¶ | pilcrow sign |
| 183 |  | · | · | · | middle dot |
| 184 |  | ¸ | ¸ | ¸ | cedilla |
| 185 |  | ¹ | ¹ | ¹ | superscript one |
| 186 |  | º | º | º | masculine ordinal indicator |
| 187 |  | » | » | » | right-pointing double angle quotation mark |
| 188 |  | ¼ | ¼ | ¼ | vulgar fraction one quarter |
| 189 |  | ½ | ½ | ½ | vulgar fraction one half |
| 190 |  | ¾ | ¾ | ¾ | vulgar fraction three quarters |
| 191 |  | ¿ | ¿ | ¿ | inverted question mark |
| 192 |  | À | À | À | Latin capital letter A with grave |
| 193 |  | Á | Á | Á | Latin capital letter A with acute |
| 194 |  | Â | Â | Â | Latin capital letter A with circumflex |
| 195 |  | Ã | Ã | Ã | Latin capital letter A with tilde |
| 196 |  | Ä | Ä | Ä | Latin capital letter A with diaeresis |
| 197 |  | Å | Å | Å | Latin capital letter A with ring above |
| 198 |  | Æ | Æ | Æ | Latin capital letter AE |
| 199 |  | Ç | Ç | Ç | Latin capital letter C with cedilla |
| 200 |  | È | È | È | Latin capital letter E with grave |
| 201 |  | É | É | É | Latin capital letter E with acute |
| 202 |  | Ê | Ê | Ê | Latin capital letter E with circumflex |
| 203 |  | Ë | Ë | Ë | Latin capital letter E with diaeresis |
| 204 |  | Ì | Ì | Ì | Latin capital letter I with grave |
| 205 |  | Í | Í | Í | Latin capital letter I with acute |
| 206 |  | Î | Î | Î | Latin capital letter I with circumflex |
| 207 |  | Ï | Ï | Ï | Latin capital letter I with diaeresis |
| 208 |  | Ð | Ð | Ð | Latin capital letter Eth |
| 209 |  | Ñ | Ñ | Ñ | Latin capital letter N with tilde |
| 210 |  | Ò | Ò | Ò | Latin capital letter O with grave |
| 211 |  | Ó | Ó | Ó | Latin capital letter O with acute |
| 212 |  | Ô | Ô | Ô | Latin capital letter O with circumflex |
| 213 |  | Õ | Õ | Õ | Latin capital letter O with tilde |
| 214 |  | Ö | Ö | Ö | Latin capital letter O with diaeresis |
| 215 |  | × | × | × | multiplication sign |
| 216 |  | Ø | Ø | Ø | Latin capital letter O with stroke |
| 217 |  | Ù | Ù | Ù | Latin capital letter U with grave |
| 218 |  | Ú | Ú | Ú | Latin capital letter U with acute |
| 219 |  | Û | Û | Û | Latin capital letter U with circumflex |
| 220 |  | Ü | Ü | Ü | Latin capital letter U with diaeresis |
| 221 |  | Ý | Ý | Ý | Latin capital letter Y with acute |
| 222 |  | Þ | Þ | Þ | Latin capital letter Thorn |
| 223 |  | ß | ß | ß | Latin small letter sharp s |
| 224 |  | à | à | à | Latin small letter a with grave |
| 225 |  | á | á | á | Latin small letter a with acute |
| 226 |  | â | â | â | Latin small letter a with circumflex |
| 227 |  | ã | ã | ã | Latin small letter a with tilde |
| 228 |  | ä | ä | ä | Latin small letter a with diaeresis |
| 229 |  | å | å | å | Latin small letter a with ring above |
| 230 |  | æ | æ | æ | Latin small letter ae |
| 231 |  | ç | ç | ç | Latin small letter c with cedilla |
| 232 |  | è | è | è | Latin small letter e with grave |
| 233 |  | é | é | é | Latin small letter e with acute |
| 234 |  | ê | ê | ê | Latin small letter e with circumflex |
| 235 |  | ë | ë | ë | Latin small letter e with diaeresis |
| 236 |  | ì | ì | ì | Latin small letter i with grave |
| 237 |  | í | í | í | Latin small letter i with acute |
| 238 |  | î | î | î | Latin small letter i with circumflex |
| 239 |  | ï | ï | ï | Latin small letter i with diaeresis |
| 240 |  | ð | ð | ð | Latin small letter eth |
| 241 |  | ñ | ñ | ñ | Latin small letter n with tilde |
| 242 |  | ò | ò | ò | Latin small letter o with grave |
| 243 |  | ó | ó | ó | Latin small letter o with acute |
| 244 |  | ô | ô | ô | Latin small letter o with circumflex |
| 245 |  | õ | õ | õ | Latin small letter o with tilde |
| 246 |  | ö | ö | ö | Latin small letter o with diaeresis |
| 247 |  | ÷ | ÷ | ÷ | division sign |
| 248 |  | ø | ø | ø | Latin small letter o with stroke |
| 249 |  | ù | ù | ù | Latin small letter u with grave |
| 250 |  | ú | ú | ú | Latin small letter u with acute |
| 251 |  | û | û | û | Latin small letter with circumflex |
| 252 |  | ü | ü | ü | Latin small letter u with diaeresis |
| 253 |  | ý | ý | ý | Latin small letter y with acute |
| 254 |  | þ | þ | þ | Latin small letter thorn |
| 255 |  | ÿ | ÿ | ÿ | Latin small letter y with diaeresis |

## The ASCII Character Set

ASCII uses the values from 0 to 31 (and 127) for control characters.

ASCII uses the values from 32 to 126 for letters, digits, and symbols.

ASCII does not use the values from 128 to 255.

## The ANSI Character Set (Windows-1252)

ANSI is identical to ASCII for the values from 0 to 127.

ANSI has a proprietary set of characters for the values from 128 to 159.

ANSI is identical to UTF-8 for the values from 160 to 255.

## The ISO-8859-1 Character Set

8859-1 is identical to ASCII for the values from 0 to 127.

8859-1 does not use the values from 128 to 159.

8859-1 is identical to UTF-8 for the values from 160 to 255.

## The UTF-8 Character Set

UTF-8 is identical to ASCII for the values from 0 to 127.

UTF-8 does not use the values from 128 to 159.

UTF-8 is identical to both ANSI and 8859-1 for the values from 160 to 255.

UTF-8 continues from the value 256 with more than 10 000 different characters.

For a closer look, study our [Complete HTML Character Set Reference](http://www.w3schools.com/charsets/default.asp).

# HTML Uniform Resource Locators

[« Previous](http://www.w3schools.com/html/html_charset.asp)

[Next Chapter »](http://www.w3schools.com/html/html_xhtml.asp)

A URL is another word for a web address.

A URL can be composed of words (w3schools.com), or an Internet Protocol (IP) address (192.68.20.50).

Most people enter the name when surfing, because names are easier to remember than numbers.

## URL - Uniform Resource Locator

Web browsers request pages from web servers by using a URL.

A Uniform Resource Locator (URL) is used to address a document (or other data) on the web.

A web address, like <http://www.w3schools.com/html/default.asp> follows these syntax rules:

scheme://prefix.domain:port/path/filename

Explanation:

* **scheme** - defines the **type** of Internet service (most common is **http or https**)
* **prefix** - defines a domain **prefix** (default for http is **www**)
* **domain** - defines the Internet **domain name**(like w3schools.com)
* **port** - defines the **port number**at the host (default for http is **80**)
* **path** - defines a **path** at the server (If omitted: the root directory of the site)
* **filename** - defines the name of a document or resource

## 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 |

## URL Encoding

URLs can only be sent over the Internet using the [ASCII character-set](http://www.w3schools.com/charsets/ref_html_ascii.asp). 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 %20.

## Try It Yourself

Top of Form

 

Bottom of Form

If you click "Submit", the browser will URL encode the input before it is sent to the server.

A page at the server will display the received input.

Try some other input and click Submit again.

## ASCII Encoding Examples

Your browser will encode input, according to the character-set used in your page.

The default character-set in HTML5 is UTF-8.

|  |  |  |
| --- | --- | --- |
| **Character** | **From Windows-1252** | **From UTF-8** |
| € | %80 | %E2%82%AC |
| £ | %A3 | %C2%A3 |
| © | %A9 | %C2%A9 |
| ® | %AE | %C2%AE |
| À | %C0 | %C3%80 |
| Á | %C1 | %C3%81 |
| Â | %C2 | %C3%82 |
| Ã | %C3 | %C3%83 |
| Ä | %C4 | %C3%84 |
| Å | %C5 | %C3%85 |

For a complete reference of all URL encodings, visit our [URL Encoding Reference](http://www.w3schools.com/tags/ref_urlencode.asp).

# HTML and XHTML

[« Previous](http://www.w3schools.com/html/html_urlencode.asp)

[Next Chapter »](http://www.w3schools.com/html/html_forms.asp)

XHTML is HTML written as XML.

## What Is XHTML?

* XHTML stands for E**X**tensible **H**yper**T**ext **M**arkup **L**anguage
* 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>

Today's market consists of different browser technologies. Some browsers run on computers, and some browsers run on mobile phones or other small devices. Smaller devices often lack the resources or power to interpret "bad" markup.

XML is a markup language where documents must be marked up correctly (be "well-formed").

If you want to study XML, please read our [XML tutorial](http://www.w3schools.com/xml/default.asp).

By combining the strengths of HTML and XML, XHTML was developed.

XHTML is HTML redesigned as XML.

## The Most Important Differences from HTML:

### Document Structure

* XHTML DOCTYPE is **mandatory**
* The xmlns attribute in <html> is **mandatory**
* <html>, <head>, <title>, and <body> are **mandatory**

### XHTML Elements

* XHTML elements must be **properly nested**
* XHTML elements must always be **closed**
* XHTML elements must be in **lowercase**
* XHTML documents must have **one root element**

### XHTML Attributes

* Attribute names must be in **lower case**
* Attribute values must be **quoted**
* Attribute minimization is **forbidden**

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

An XHTML document must have an XHTML DOCTYPE declaration.

A complete list of all the [XHTML Doctypes](http://www.w3schools.com/tags/tag_doctype.asp) is found in our HTML Tags Reference.

The <html>, <head>, <title>, and <body> elements must also be present, and the xmlns attribute in <html> must specify the xml namespace for the document.

This example shows an XHTML document with a minimum of required tags:

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"  
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">  
  
<html xmlns="http://www.w3.org/1999/xhtml">  
  
<head>  
  <title>Title of document</title>  
</head>  
  
<body>  
  some content   
</body>  
  
</html>

## 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>

In XHTML, all elements must be properly nested within each other, like this:

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

## XHTML Elements Must Always Be Closed

This is wrong:

<p>This is a paragraph  
<p>This is another paragraph

This is correct:

<p>This is a paragraph</p>  
<p>This is another paragraph</p>

## Empty Elements Must Also Be Closed

This is wrong:

A break: <br>  
A horizontal rule: <hr>  
An image: <img src="happy.gif" alt="Happy face">

This is correct:

A break: <br />  
A horizontal rule: <hr />  
An image: <img src="happy.gif" alt="Happy face" />

## XHTML Elements Must Be In Lower Case

This is wrong:

<BODY>  
<P>This is a paragraph</P>  
</BODY>

This is correct:

<body>  
<p>This is a paragraph</p>  
</body>

## XHTML Attribute Names Must Be In Lower Case

This is wrong:

<table WIDTH="100%">

This is correct:

<table width="100%">

## Attribute Values Must Be Quoted

This is wrong:

<table width=100%>

This is correct:

<table width="100%">

## Attribute Minimization Is Forbidden

Wrong:

<input type="checkbox" name="vehicle" value="car" checked />

Correct:

<input type="checkbox" name="vehicle" value="car" checked="checked" />

Wrong:

<input type="text" name="lastname" disabled />

Correct:

<input type="text" name="lastname" disabled="disabled" />

## How to Convert from HTML to XHTML

1. Add an XHTML <!DOCTYPE> to the first line of every page
2. Add an xmlns attribute to the html element of every page
3. Change all element names to lowercase
4. Close all empty elements
5. Change all attribute names to lowercase
6. Quote all attribute values

## Validate XHTML With The W3C Validator

Top of Form

Put your web address in the box below:





## HTML Layout Elements

Websites often display content in multiple columns (like a magazine or newspaper).

HTML5 offers new semantic elements that define the different parts of a web page:

|  |  |
| --- | --- |
| HTML5 Semantic Elements | * <header> - Defines a header for a document or a section * <nav> - Defines a container for navigation links * <section> - Defines a section in a document * <article> - Defines an independent self-contained article * <aside> - Defines content aside from the content (like a sidebar) * <footer> - Defines a footer for a document or a section * <details> - Defines additional details * <summary> - Defines a heading for the <details> element |

## HTML Layout Techniques

There are four different ways to create multicolumn layouts. Each way has its pros and cons:

* HTML tables
* CSS float property
* CSS framework
* CSS flexbox

## Which One to Choose?

### HTML Tables

The <table> element was not designed to be a layout tool! The purpose of the <table> element is to display tabular data. So, do not use tables for your page layout! They will bring a mess into your code. And imagine how hard it will be to redesign your site after a couple of months.

**Tip:** Do NOT use tables for your page layout!

### CSS Frameworks

If you want to create your layout fast, you can use a framework, like [W3.CSS](http://www.w3schools.com/w3css/default.asp) or [Bootstrap](http://www.w3schools.com/bootstrap/default.asp).

### CSS Floats

It is common to do entire web layouts using the CSS float property. Float is easy to learn - you just need to remember how the float and clear properties work. Disadvantages: Floating elements are tied to the document flow, which may harm the flexibility. Learn more about float in our [CSS Float and Clear](http://www.w3schools.com/css/css_float.asp) chapter.

### CSS Flexbox

Flexbox is a new layout mode in CSS3.

Use of flexbox ensures that elements behave predictably when the page layout must accommodate different screen sizes and different display devices. Disadvantages: Does not work in IE10 and earlier.

Learn more about flexbox in our [CSS Flexbox](http://www.w3schools.com/css/css3_flexbox.asp) chapter.

Bottom of Form