

Character Entity References

Characters not found in the normal alphanumeric character set, such as < and &, must be specified in HTML and XHTML documents using character references. This is known as *escaping* the character. Using the standard desktop publishing keyboard commands (such as Option-G for the © symbol) within an HTML document will not produce the desired character when the document is rendered in a browser. In fact, the browser generally displays the numeric entity for the character.

In (X)HTML documents, escaped characters are indicated by character references that begin with & and end with ;. The character may be referred to by its Numeric Character Reference (NCR) or a predefined character entity name.

A Numeric Character Reference refers to a character by its Unicode code point in either decimal or hexadecimal form (for more information on Unicode and code points, see Chapter 6). Decimal character references use the syntax &#nnnn;. Hexadecimal values are indicated by an “x”: &#xhhhh;. For example, the less-than (<) character could be identified as < (decimal) or < (hexadecimal).

Character entities are abbreviated names for characters, such as < for the less-than symbol. Character entities are predefined in the DTDs of markup languages such as HTML and XHTML as a convenience to authors, because they may be easier to remember than Numeric Character References.



XHTML includes the XML entity declaration for the apostrophe ('). In HTML, the apostrophe character entity was curiously omitted, so its numeric reference (') must be used instead.

Table 10-3 presents the (X)HTML character entities and numeric character references for commonly used special characters. The complete list of character entities defined in HTML 4.01 and XHTML 1.0/1.1 appears in Appendix C.

Table 10-3. Common special characters and their character entities

Character	Description	Entity	Decimal	Hex
	Character space (nonbreaking space)	 	 	
&	Ampersand	&	&	&
<	Less-than sign (useful for displaying markup on a web page)	<	<	<
>	Greater-than sign (useful for displaying markup on a web page)	>	>	>
'	Apostrophe	' (XHTML only)	'	'
“	Left curly quotes	&lquo;	“	“
”	Right curly quotes	&rquo;	”	”
™	Trademark	™	™	™
£	Pound symbol	£	£	£