

Allgemeines zur HTML-Zeichenreferenz

In der Zeichenreferenz geht es um Zeichen, die zwischen dem Anfangs- und dem End-Tag eines HTML-Elements als Text stehen können (`#PCDATA`, *parsed character data*), sowie um Zeichen, die bei Wertzuweisungen an viele Attribute notiert werden können (`#CDATA`, *character data*). Bei einem HTML-Konstrukt wie `<p class="Aussage">Zeichen bedeuten viel</p>` ist zum einen die Zeichenfolge `Zeichen bedeuten viel` relevant, und zum anderen die Zeichenfolge `Aussage`.

Seit HTML 4.0 kann jedes Zeichen aus dem Zeichenvorrat nach ISO 10646 (= Unicode-Standard) notiert werden. Beliebige Zeichen, aus diesem großen Zeichenvorrat können durch eine spezielle numerische Notation erzeugt werden. Das Notations-Schema für solche Notationen lautet:

`&#[x][Nummer];`

Beispiele einer Notation in HTML: `α` `α`

Das `x` notieren Sie dann, wenn Sie die Nummer hexadezimal angeben. Bei dezimaler Angabe lassen Sie das `x` weg.

Neben der numerischen Unicode-Notation gibt es in HTML für häufiger verwendete Sonderzeichen auch so genannte **named entities**, benannte Zeichen. Alle benannten Zeichen werden nach dem folgenden Schema notiert:

`&[Name];`

Beispiel einer Notation in HTML: `α`

Deutsche Umlaute und scharfes S / Euro-Zeichen

Wenn Sie in Ihrer HTML-Datei sonst keine Angaben zum verwendeten Zeichensatz machen, sollten Sie im Hinblick auf die internationale Verwendung deutsche Umlaute und scharfes S durch spezielle, dafür vorgesehene **benannte Zeichen** ersetzen. Das gilt für den gesamten Inhalt einer HTML-Datei.

Ersetzen Sie das Zeichen `ä` durch die Zeichenfolge `ä`

Ersetzen Sie das Zeichen `Ä` durch die Zeichenfolge `Ä`

Ersetzen Sie das Zeichen `ö` durch die Zeichenfolge `ö`

Ersetzen Sie das Zeichen `Ö` durch die Zeichenfolge `Ö`

Ersetzen Sie das Zeichen `ü` durch die Zeichenfolge `ü`

Ersetzen Sie das Zeichen `Û` durch die Zeichenfolge `Ü`

Ersetzen Sie das Zeichen `ß` durch die Zeichenfolge `ß`

Ersetzen Sie das Zeichen `€` durch die Zeichenfolge `€`

```
In M&uuml;nchen steht ein Hofbr&auml;uhaus.  
Dort gibt es Bier aus Ma&szlig;kr&uuml;gen, eine Ma&szlig; kostet 7 &euro;
```

HTML-eigene Zeichen maskieren

Wenn in Ihrem Text Zeichen vorkommen, die in HTML eine bestimmte Bedeutung haben, müssen Sie diese Zeichen **maskieren**. Die folgenden Zeichen müssen Sie wie folgt maskieren:

Ersetzen Sie das Zeichen `<` durch die Zeichenfolge `<`

Ersetzen Sie das Zeichen `>` durch die Zeichenfolge `>`

Ersetzen Sie das Zeichen `&` durch die Zeichenfolge `&`

Ersetzen Sie das Zeichen `"` durch die Zeichenfolge `"`

```
Das ist ein &lt;HTML-Tag&gt; und das eine GmbH &amp; Co. KG  
&quot;Text steht in Anf&uuml;hrungszeichen&quot;
```

Die ISO-8859-Zeichensätze

Die ISO-8859-Familie wurde vom [European Computer Manufacturer's Association \(ECMA\)](#) entwickelt. Es handelt sich um ein Set von standardisierten Zeichensätzen für alphabetische Schriften. Dazu gehören die lateinischen Schriften, auf denen die meisten Sprachen Westeuropas und Amerikas beruhen, oder etwa die kyrillischen Schriften.

Alle Zeichensätze dieser Familie basieren auf der 1-Byte-Abbildung eines Zeichens. Das heißt, alle Zeichensätze enthalten 256 mögliche Zeichen. Bei allen Zeichensätzen sind die ersten 128 Zeichen, also die Zeichen mit den Werten 0 bis 127, identisch mit dem [ASCII-Zeichensatz](#). Das hat den Vorteil, dass die üblichen lateinischen Groß- und Kleinbuchstaben, die arabischen Ziffern und die üblichen Sonderzeichen wie Satzzeichen oder kaufmännische Zeichen in all diesen Zeichensätzen immer zur Verfügung stehen.

ASCII-Zeichensatz										
+	0	1	2	3	4	5	6	7	8	9
30				!	"	#	\$	%	&	'
40	()	*	+	,	-	.	/	0	1
50	2	3	4	5	6	7	8	9	:	;
60	<	=	>	?	@	A	B	C	D	E
70	F	G	H	I	J	K	L	M	N	O
80	P	Q	R	S	T	U	V	W	X	Y
90	Z	[\]	^	_	`	a	b	c
100	d	e	f	g	h	i	j	k	l	m
110	n	o	p	q	r	s	t	u	v	w
120	x	y	z	{		}	~			

Links ist der ASCII-Zeichensatz abgebildet. Die darstellbaren Zeichen dieses Zeichensatzes haben die Werte zwischen 32 und 127. Wert 32 ist das Leerzeichen. Alle ISO-8859-Zeichensätze enthalten diesen Zeichensatz.

Den Wert eines Zeichens ermitteln Sie, indem Sie die Werte aus Spalten- und Zeilenüberschrift des Zeichens addieren. Ein großes K hat in dieser Tabelle beispielsweise den Zeilenwert 70 und den Spaltenwert 5. Daraus ergibt sich der Wert 75 für dieses Zeichen. Zeichenwerte können bei der Verwendung von Zeichensätzen und Unicodes in HTML eine Rolle spielen.

Die ersten 32 Zeichen (Zeichenwerte 0 bis 31) sind für Steuerzeichen reserviert, zum Beispiel für Zeilenumbrüche.

iso-8859-1										
+	0	1	2	3	4	5	6	7	8	9
160		ı	ø	£	¤	¥	ı	š	ˆ	ø
170	æ	«	¬	-	ø	-	°	±	²	³
180	˘	µ	¶	.	˙	ı	º	»	¼	½
190	¾	¿	À	Á	Â	Ã	Ä	Å	Æ	Ç
200	È	É	Ê	Ë	Ì	Í	Î	Ï	Ð	Ñ
210	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û
220	Ü	Ý	Þ	ß	à	á	â	ã	ä	å
230	æ	ç	è	é	ê	ë	ì	í	î	ï
240	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù
250	ú	û	ü	ý	þ	ÿ				

ISO-8859-1 (Latin-1)

Dieser Zeichensatz enthält die schriftspezifischen Zeichen für westeuropäische und amerikanische Sprachen. Der Zeichensatz deckt die Sprachen Albanisch, Dänisch, Deutsch, Englisch, Farörisch, Finnisch, Französisch, Galizisch, Irisch, Isländisch, Italienisch, Katalanisch, Niederländisch, Norwegisch, Portugiesisch, Schwedisch und Spanisch ab. Lediglich einzelne Zeichen wie das niederländische "ij" oder die deutschen Anführungszeichen unten fehlen.

Das Zeichen mit dem Wert 160 ist in allen ISO-8859-Zeichensätzen ein erzwungenes Leerzeichen.

Benannte Zeichen für den Zeichensatz ISO 8859-1

Die HTML-Namen dieser Zeichen sind seit HTML 3.2 verfügbar.

Zeichen	Beschreibung	Name in HTML	Unicode in HTML
	Erzwungenes Leerzeichen	 	
¡	umgekehrtes Ausrufezeichen	¡	¡
¢	Cent-Zeichen	¢	¢
£	Pfund-Zeichen	£	£
¤	Währungs-Zeichen	¤t;	¤
¥	Yen-Zeichen	¥	¥
¦	durchbrochener Strich	¦	¦
§	Paragraph-Zeichen	§	§
¨	Pünktchen oben	¨	¨
©	Copyright-Zeichen	©	©
ª	Ordinal-Zeichen weiblich	ª	ª
«	angewinkelte Anführungszeichen links	«	«
¬	Verneinungs-Zeichen	¬	¬
–	kurzer Trennstrich	­	­
®	Registriermarke-Zeichen	®	®
¯	Überstrich	¯	¯
°	Grad-Zeichen	°	°
±	Plusminus-Zeichen	±	±
²	Hoch-2-Zeichen	²	²
³	Hoch-3-Zeichen	³	³
´	Acute-Zeichen	´	´
µ	Mikro-Zeichen	µ	µ
¶	Absatz-Zeichen	¶	¶
·	Mittelpunkt	·	·
¸	Häkchen unten	¸	¸
¹	Hoch-1-Zeichen	¹	¹
º	Ordinal-Zeichen männlich	º	º
»	angewinkelte Anführungszeichen rechts	»	»
¼	ein Viertel	¼	¼
½	ein Halb	½	½
¾	drei Viertel	¾	¾
¿	umgekehrtes Fragezeichen	¿	¿
À	A mit Accent grave	À	À
Á	A mit Accent acute	Á	Á
Â	A mit Circumflex	Â	Â
Ã	A mit Tilde	Ã	Ã
Ä	A Umlaut	Ä	Ä
Å	A mit Ring	Å	Å
Æ	A mit legiertem E	Æ	Æ
Ç	C mit Häkchen	Ç	Ç
È	E mit Accent grave	È	È
É	E mit Accent acute	É	É
Ê	E mit Circumflex	Ê	Ê
Ë	E Umlaut	Ë	Ë
Ì	I mit Accent grave	Ì	Ì
Í	I mit Accent acute	Í	Í
Î	I mit Circumflex	Î	Î
Ï	I Umlaut	Ï	Ï
Ð	Eth (isländisch)	Ð	Ð

Ñ	N mit Tilde	Ñ	Ñ
Ò	O mit Accent grave	Ò	Ò
Ó	O mit Accent acute	Ó	Ó
Ô	O mit Circumflex	Ô	Ô
Õ	O mit Tilde	Õ	Õ
Ö	O Umlaut	Ö	Ö
×	Mal-Zeichen	×	×
Ø	O mit Schrägstrich	Ø	Ø
Ù	U mit Accent grave	Ù	Ù
Ú	U mit Accent acute	Ú	Ú
Û	U mit Circumflex	Û	Û
Ü	U Umlaut	Ü	Ü
Ý	Y mit Accent acute	Ý	Ý
Þ	THORN (isländisch)	Þ	Þ
ß	scharfes S	ß	ß
à	a mit Accent grave	à	à
á	a mit Accent acute	á	á
â	a mit Circumflex	â	â
ã	a mit Tilde	ã	ã
ä	a Umlaut	ä	ä
å	a mit Ring	å	å
æ	a mit legiertem e	æ	æ
ç	c mit Häkchen	ç	ç
è	e mit Accent grave	è	è
é	e mit Accent acute	é	é
ê	e mit Circumflex	ê	ê
ë	e Umlaut	ë	ë
ì	i mit Accent grave	ì	ì
í	i mit Accent acute	í	í
î	i mit Circumflex	î	î
ï	i Umlaut	ï	ï
ð	eth (isländisch)	ð	ð
ñ	n mit Tilde	ñ	ñ
ò	o mit Accent grave	ò	ò
ó	o mit Accent acute	ó	ó
ô	o mit Circumflex	ô	ô
õ	o mit Tilde	õ	õ
ö	o Umlaut	ö	ö
÷	Divisions-Zeichen	÷	÷
ø	o mit Schrägstrich	ø	ø
ù	u mit Accent grave	ù	ù
ú	u mit Accent acute	ú	ú
û	u mit Circumflex	û	û
ü	u Umlaut	ü	ü
ý	y mit Accent acute	ý	ý
þ	thorn (isländisch)	þ	þ
ÿ	y Umlaut	ÿ	ÿ