

HTML5 - URL ENCODING

http://www.tutorialspoint.com/html5/html5_url_encoding.htm

Copyright © tutorialspoint.com

URL encoding is the practice of translating unprintable characters or characters with special meaning within URLs to a representation that is unambiguous and universally accepted by web browsers and servers. These characters include –

- **ASCII control characters** – Unprintable characters typically used for output control. Character ranges 00-1F hex 0 – 31~~decimal~~ and 7F 127~~decimal~~. A complete encoding table is given below.
- **Non-ASCII control characters** – These are characters beyond the ASCII character set of 128 characters. This range is part of the ISO-Latin character set and includes the entire "top half" of the ISO-Latin set 80-FF hex 128 – 255~~decimal~~. A complete encoding table is given below.
- **Reserved characters** – These are special characters such as the dollar sign, ampersand, plus, common, forward slash, colon, semi-colon, equals sign, question mark, and "at" symbol. All of these can have different meanings inside a URL so need to be encoded. A complete encoding table is given below.
- **Unsafe characters** – These are space, quotation marks, less than symbol, greater than symbol, pound character, percent character, Left Curly Brace, Right Curly Brace, Pipe, Backslash, Caret, Tilde, Left Square Bracket, Right Square Bracket, Grave Accent. These characters present the possibility of being misunderstood within URLs for various reasons. These characters should also always be encoded. A complete encoding table is given below.

The encoding notation replaces the desired character with three characters: a percent sign and two hexadecimal digits whose correspond to the position of the character in the ASCII character set.

Example

One of the most common special characters is the space. You can't type a space in a URL directly. A space position in the character set is 20 hexadecimal. So you can use %20 in place a space when passing your request to the server.

```
http://www.example.com/new%20pricing.html
```

This URL actually retrieves a document named *new pricing.html* from the www.example.com

ASCII control characters encoding

This includes the encoding for character ranges 00-1F hex 0 – 31~~decimal~~ and 7F 127~~decimal~~

Decimal	Hex Value	Character	URL Encode
0	00		%00
1	01		%01
2	02		%02
3	03		%03
4	04		%04
5	05		%05
6	06		%06
7	07		%07

8	08	backspace	%08
9	09	tab	%09
10	0a	linefeed	%0a
11	0b		%0b
12	0c		%0c
13	0d	carriage return	%0d
14	0e		%0e
15	0f		%0f
16	10		%10
17	11		%11
18	12		%12
19	13		%13
20	14		%14
21	15		%15
22	16		%16
23	17		%17
24	18		%18
25	19		%19
26	1a		%1a
27	1b		%1b
28	1c		%1c
29	1d		%1d
30	1e		%1e
31	1f		%1f
127	7f		%7f

Non-ASCII control characters encoding

This includes the encoding for the entire "top half" of the ISO-Latin set 80-FF hex 128 – 255*decimal*.

Decimal	Hex Value	Character	URL Encode
128	80	€	%80
129	81		%81
130	82	,	%82
131	83	f	%83
132	84	„	%84

133	85	...	%85
134	86	†	%86
135	87	‡	%87
136	88	^	%88
137	89	‰	%89
138	8a	Š	%8a
139	8b	<	%8b
140	8c	Œ	%8c
141	8d		%8d
142	8e	Ž	%8e
143	8f		%8f
144	90		%90
145	91	‘	%91
146	92	’	%92
147	93	“	%93
148	94	”	%94
149	95	•	%95
150	96	–	%96
151	97	—	%97
152	98	~	%98
153	99	™	%99
154	9a	š	%9a
155	9b	>	%9b
156	9c	œ	%9c
157	9d		%9d
158	9e	ž	%9e
159	9f	Ÿ	%9f
160	a0		%a0
161	a1	ı	%a1
162	a2	¢	%a2
163	a3	£	%a3
164	a4	¤	%a4
165	a5	¥	%a5
166	a6	¦	%a6
167	a7	§	%a7

168	a8	¨	%a8
169	a9	©	%a9
170	aa	ª	%aa
171	ab	«	%ab
172	ac	¬	%ac
173	ad		%ad
174	ae	®	%ae
175	af	-	%af
176	b0	º	%b0
177	b1	±	%b1
178	b2	²	%b2
179	b3	³	%b3
180	b4	´	%b4
181	b5	µ	%b5
182	b6	¶	%b6
183	b7	·	%b7
184	b8	¸	%b8
185	b9	¹	%b9
186	ba	º	%ba
187	bb	»	%bb
188	bc	¼	%bc
189	bd	½	%bd
190	be	¾	%be
191	bf	¿	%bf
192	c0	À	%c0
193	c1	Á	%c1
194	c2	Â	%c2
195	c3	Ã	%c3
196	c4	Ä	%c4
197	c5	Å	%c5
198	c6	Æ	%v6
199	c7	Ç	%c7
200	c8	È	%c8
201	c9	É	%c9
202	ca	Ê	%ca

203	cb	Ě	%cb
204	cc	ì	%cc
205	cd	í	%cd
206	ce	î	%ce
207	cf	ï	%cf
208	d0	Đ	%d0
209	d1	Ñ	%d1
210	d2	Ò	%d2
211	d3	Ó	%d3
212	d4	Ô	%d4
213	d5	Õ	%d5
214	d6	Ö	%d6
215	d7	×	%d7
216	d8	Ø	%d8
217	d9	Ù	%d9
218	da	Ú	%da
219	db	Û	%db
220	dc	Ü	%dc
221	dd	Ý	%dd
222	de	Ɔ	%de
223	df	ß	%df
224	e0	à	%e0
225	e1	á	%e1
226	e2	â	%e2
227	e3	ã	%e3
228	e4	ä	%e4
229	e5	å	%e5
230	e6	æ	%e6
231	e7	ç	%e7
232	e8	è	%e8
233	e9	é	%e9
234	ea	ê	%ea
235	eb	ë	%eb
236	ec	ì	%ec
237	ed	í	%ed

238	ee	î	%ee
239	ef	ï	%ef
240	f0	ð	%f0
241	f1	ñ	%f1
242	f2	ò	%f2
243	f3	ó	%f3
244	f4	ô	%f4
245	f5	õ	%f5
246	f6	ö	%f6
247	f7	÷	%f7
248	f8	ø	%f8
249	f9	ù	%f9
250	fa	ú	%fa
251	fb	û	%fb
252	fc	ü	%fc
253	fd	ý	%fd
254	fe	þ	%fe
255	ff	ÿ	%ff

Reserved characters encoding

Following is the table to be used to encode reserved characters.

Decimal	Hex Value	Char	URL Encode
36	24	\$	%24
38	26	&	%26
43	2b	+	%2b
44	2c	,	%2c
47	2f	/	%2f
58	3a	:	%3a
59	3b	;	%3b
61	3d	=	%3d
63	3f	?	%3f
64	40	@	%40

Unsafe characters encoding

Following is the table to be used to encode unsafe characters.

Decimal	Hex Value	Char	URL Encode
32	20	space	%20
34	22	"	%22
60	3c	<	%3c
62	3e	>	%3e
35	23	#	%23
37	25	%	%25
123	7b	{	%7b
125	7d	}	%7d
124	7c		%7c
92	5c	\	%5c
94	5e	^	%5e
126	7e	~	%7e
91	5b	[	%5b
93	5d	]	%5d
96	60	`	%60