



Università degli Studi di Padova
Facoltà di Lettere e Filosofia

Progettazione di siti web (ING-INF/05)

LMSGC, mutuato da LMLIN e Sc. Stat. – a.a. 2010-11

Prof. Maristella Agosti, Ing. Marco Dussin

La codifica di carattere

Indice:

- ↪ La codifica di un oggetto digitale
- ↪ La codifica di carattere



La codifica di un oggetto digitale

- In informatica la codifica viene utilizzata per trasformare gli oggetti digitali in codici numerici, che possano essere memorizzati o trasmessi.
- Ogni oggetto digitale viene codificato: documenti di testo, immagini, documenti multimediali come video o audio, programmi e procedure, informazioni da trasmettere, streaming audio o video...



La codifica di carattere

- La **codifica di carattere** (o *charset*) consiste in un codice che permette l'associazione tra un insieme di caratteri (tipicamente i *grafemi* di un alfabeto) ed un insieme di altri oggetti, per facilitare la memorizzazione o la trasmissione di tali caratteri. La corrispondenza è chiamata “tabella caratteri”.

INTERNATIONAL MORSE CODE

1. A dash is equal to three dots.
2. The space between parts of the same letter is equal to one dot.
3. The space between two letters is equal to three dots.
4. The space between two words is equal to five dots.

A	• —	U	• • —
B	— • • •	V	• • • —
C	— • — •	W	— • —
D	— • • •	X	— • • —
E	•	Y	— • — —
F	• • — •	Z	— — • •
G	— — •		
H	• • • •		
I	• •		
J	• — — —		
K	— • —	1	• — — — —
L	• — • •	2	• • — — —
M	— — •	3	• • • — —
N	— •	4	• • • • —
O	— — —	5	• • • • •
P	• — — •	6	— • • • •
Q	— — • —	7	— — • • •
R	• — • •	8	— — — • •
S	• • •	9	— — — — •
T	—	0	— — — — —

A		Alfa	N		November
B		Bravo	O		Oscar
C		Charlie	P		Papa
D		Delta	Q		Quebec
E		Echo	R		Romeo
F		Foxtrot	S		Sierra
G		Golf	T		Tango
H		Hotel	U		Uniform
I		India	V		Victor
J		Juliet	W		Whiskey
K		Kilo	X		X-ray
L		Lima	Y		Yankee
M		Mike	Z		Zulu

0		Intelligenza
1		1a Ripetitrice
2		2a Ripetitrice
3		3a Ripetitrice
4		3a Ripetitrice
5		3a Ripetitrice
6		3a Ripetitrice
7		3a Ripetitrice
8		3a Ripetitrice
9		3a Ripetitrice



La codifica di carattere

- La prima codifica di carattere ad essere standardizzata nel mondo dell'informatica è la codifica ASCII (definita dall'ANSI, American National Standards Institute), risalente agli anni 60. Essa rappresenta i caratteri alfanumerici utilizzando per ognuno 7 bit, per un totale di 128 possibili caratteri rappresentabili (lettere maiuscole e minuscole, numeri, segni di interpunzione, caratteri di controllo).
- E' evidente che la codifica ASCII da sola non è sufficiente a rappresentare tutti i grafemi del mondo: esistono quindi delle sue estensioni. Nascono così le diverse tabelle ISO 8859, standardizzate in 14 varianti.
- La variante ISO 8859-1, detta Latin-1, è quella solitamente utilizzata per rappresentare le lettere accentate più comuni in Italia, Francia e Spagna. Il server users.lettere.unipd.it è impostato per restituire sempre le pagine html utilizzando tale codifica.
- I produttori di sistemi operativi hanno poi aggiunto loro varianti a queste tabelle, ad esempio Microsoft utilizza la sua variante di Latin-1: Windows-1252



La codifica di carattere

ISO 8859:

- ISO 8859-1 Europa occidentale
- ISO 8859-2 Europa centrale e occidentale
- ISO 8859-3 Europa centrale e meridionale
- ISO 8859-4 Europa occidentale e nazioni baltiche
- ISO 8859-5 Alfabeto cirillico
- ISO 8859-6 Alfabeto arabo
- ISO 8859-7 Alfabeto greco
- ISO 8859-8 Alfabeto ebreo
- ISO 8859-9 Europa occidentale e caratteri turchi
- ISO 8859-10 Europa occidentale e caratteri nordici
- ISO 8859-11 Alfabeto thailandese
- ISO 8859-13 Lingue baltiche e polacco
- ISO 8859-14 Lingue celtiche
- ISO 8859-15 ISO 8859-1 con aggiunta di caratteri(come il simbolo dell'euro)
- ISO 8859-16 Europa centrale(Italia, Germania ed est europeo)



La codifica di carattere

- Questa babele di caratteri e loro codifica in diverse tabelle viene risolta da Unicode, un ente che si è posto il problema di descrivere in un unico spazio concettuale tutti i caratteri esistenti in tutti gli alfabeti noti, predisponendo lo spazio anche per eventuali simboli futuri e futuribili.
- Parliamo di spazio concettuale in quanto Unicode partiziona questa che sarebbe una tabella molto grande (più di un milione di caratteri) in mappe da 64000 caratteri circa, fornendo delle codifiche standard che permettano di specificare in quale mappa debba essere cercato il carattere corrispondente alla codifica. E' in qualche modo un funzionamento simile al modo in cui vengono specificati gli indirizzi delle abitazioni: via e poi numero civico.
- Le codifiche più usate per questo standard sono la UTF-8 e la UTF-16. Esse, a differenza di ASCII, utilizzano sequenze di lunghezza variabile per codificare i caratteri.