

CODIFICA DEI CARATTERI

- **CODIFICA ASCII E ASCII ESTESO**

Sebbene la sequenza di bit dei 95 caratteri ASCII stampabili (tutte le lettere occidentali non accentate) siano sufficienti per lo scambio di informazioni usando la lingua inglese, molte altre lingue che usano l'alfabeto latino hanno bisogno di simboli addizionali non compresi nell'ASCII .

L'ISO 8859 cercò di rimediare a questo problema utilizzando l'ottavo bit del byte, permettendo la definizione di altri 128 caratteri.

Questo bit era usato in precedenza per informazioni del protocollo di trasmissione dati oppure era inutilizzato.

Comunque, erano necessari più caratteri di quanti potessero stare nella codifica a 8 bit, così furono sviluppate diverse mappe di caratteri, di cui una decina solo per coprire le lingue latine.

- **CODIFICA UNICODE**

Lo standard che sta prendendo piede e che dovrebbe essere il successore di ASCII è UTF-8, specie da quando è diventato la codifica principale di Unicode.

Questa codifica di Unicode che ormai è universalmente accettata da ogni nuovo programma offre caratteristiche interessanti grazie al fatto di sfruttare il bit di controllo del vecchio ASCII in modo più efficiente.

UTF-8 in un certo senso ha reso molto più longevo l'ASCII, inglobandolo in una codifica più adatta alle esigenze attuali.

Infatti le comunicazioni sono in centinaia di lingue, ed inoltre diverse lingue possono convivere in uno stesso documento, o in uno stesso programma: si pensi ai software di messaggistica istantanea che possono contenere contatti da varie parti del mondo dei quali visualizzare correttamente i nomi.

UTF-8 infatti si distingue dalle altre codifiche Unicode perché sfrutta il vecchio bit di parità di ASCII, non come bit di controllo, bensì come indicatore: analizza ogni byte, e se al posto del vecchio bit di parità c'è 0, allora il byte sarà letto come ASCII a 7 bit e teoricamente compatibile anche da programmi obsoleti; se però il byte corrente inizia con 1, allora sarà concatenato al byte successivo (o ai successivi).

In tal modo riesce ad includere tutti gli alfabeti delle lingue viventi, di alcune morte e potenzialmente può essere esteso per rappresentarne ancora altri (infatti spesso viene aggiornato).