

Corso di Calcolatori Elettronici I

Codifica di informazione di tipo carattere: cenni

Prof. Roberto Canonico



Università degli Studi di Napoli Federico II
Dipartimento di Ingegneria Elettrica e
delle Tecnologie dell'Informazione
Corso di Laurea in Ingegneria Informatica
Corso di Laurea in Ingegneria dell'Automazione

Rappresentazione dei caratteri

- E' necessario poter rappresentare dati di tipo carattere
 - Il tipo carattere contiene non solo le lettere dell'alfabeto (quale?!?) ma anche altri caratteri: cifre decimali, segni di interpunzione, caratteri speciali, di controllo etc.
 - Qual è la cardinalità del tipo carattere?
 - Forse è meglio chiedersi quanti bit vogliamo utilizzare per rappresentare il tipo carattere...
-

Rappresentazione dei caratteri

- D=caratteri, R=cifre binarie
 - Se $m=7$, allora $|D|=2^7=128$
 - Se $m=8$, allora $|D|=2^8=256$

 - Codici standard sono stati proposti allo scopo di facilitare lo scambio di testi codificati tra diversi sistemi di calcolo, ad esempio: ASCII, UNICODE, etc...
-

Codice ASCII

- ASCII=*American Standard Code for Information Interchange*
 - Nella sua versione originale prevede di rappresentare 128 caratteri con stringhe di 7 bit
 - I caratteri sono elencati seguendo un particolare ordine e associati a stringhe di bit corrispondenti a numeri naturali crescenti da 0 a 127 (se interpretate come numeri binari)
 - Poiché i registri di un calcolatore moderno contengono un numero di bit che è un multiplo di 8, per rispettare il codice ASCII si pone l'ottavo bit (quello più a sinistra) pari a 0
-

Codice ASCII

- Schema (notare che in questo modo si introduce un ordinamento relativo):
 - da 0 a 31: “caratteri” di controllo
 - da 32 a 47: interpunzione e caratteri speciali
 - da 48 a 57: cifre decimali
 - da 58 a 64: interpunzione e caratteri speciali
 - da 65 a 90: lettere maiuscole dell’alfabeto inglese
 - da 91 a 96: interpunzione e caratteri speciali
 - da 97 a 122: lettere minuscole alfabeto inglese
 - da 123 a 127: caratteri speciali
 - $\text{ASCII}('a') - \text{ASCII}('A') = \dots = \text{ASCII}('z') - \text{ASCII}('Z') = 32$
-

Estensioni del codice ASCII

- Utilizzando l’ottavo bit
 - i caratteri da 128 a 255 rappresentano vari caratteri speciali, simboli matematici e lettere non appartenenti all’alfabeto inglese
 - Questi codici non sono altrettanto universali
 - Non sempre sono correttamente interpretati
-

Standard UNICODE

- E' uno standard recente
 - $|D|=2^{16}$ (ma anche più di 16 bit)
 - Risolve in maniera completa il problema della specificità delle lingue e dei simboli matematici
 - E' compatibile con ASCII (lo contiene)
-