

Capitolo 3

La rappresentazione dei dati per le scienze umane

Mauro Cadei



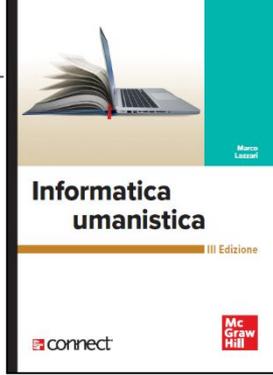
La rappresentazione dei numeri (naturali)

- rappresentazioni non posizionali
 - sistema di numerazione romano: additivo
- rappresentazioni posizionali
 - notazione decimale posizionale
 - notazione posizionale ottale
 - notazione posizionale esadecimale
 - notazione posizionale in altra base
 - notazione posizionale binaria



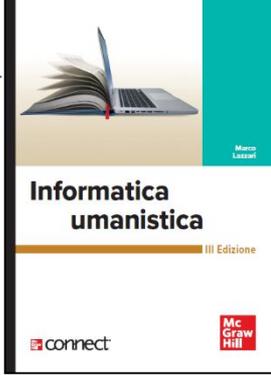
Codifica binaria dei naturali

- cifre (bit): {"0", "1"}
- esempio: $100101012 = 1 * 2^7 + 0 * 2^6 + 0 * 2^5 + 1 * 2^4 + 0 * 2^3 + 1 * 2^2 + 0 * 2^1 + 1 * 2^0 = 14910$
- con n bit si possono rappresentare 2^n numeri naturali (da 0 a 2^n-1)
- con un byte (8 bit): 256 naturali (00000000 ... 11111111)
- con un nibble (4 bit): 16 naturali (0000 ... 1111)
 - un nibble può essere rappresentato da una cifra esadecimale



Incertezza e informazione

- teoria dell'informazione: Claude Shannon (USA 1916-2001)
- prima di accedere al supporto contenente la codifica dell'informazione vi è incertezza
- l'accesso al supporto diminuisce (annulla) tale incertezza
- l'accesso al supporto porta informazione
- la differenza tra i due valori d'incertezza (probabilità) dopo e prima dell'accesso è una misura della quantità di informazione portata dal supporto
- meno è probabile una configurazione del supporto, maggiore è l'informazione che essa porta quando si presenta
- $I = \log(P_{\text{dopo}}) - \log(P_{\text{prima}})$
- se i logaritmi sono in base 2, l'unità di misura della quantità d'informazione è definita bit



Codifica analogica e digitale

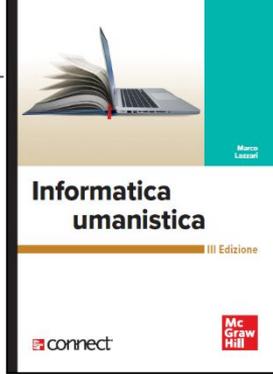
Codifica analogica

- grandezza analoga
- ad ogni incremento / decremento della grandezza corrisponde un incremento / decremento della grandezza analoga
- meta-informazione esplicita nel supporto
 - ordinale (relazione d'ordine tra i valori)
 - metrica (quantificazione delle differenze tra valori)

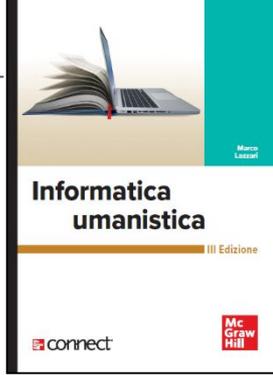
Codifica digitale

- alfabeto di simboli
- regole di codifica: specificano come ogni valore della grandezza è rappresentato da una sequenza di simboli
- più compatta

Codifica di grandezze continue

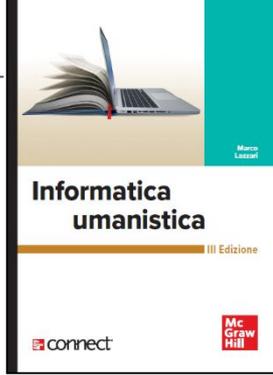


- digitalizzazione
 - quantizzazione
 - campionamento
- livelli di quantizzazione e frequenza di campionamento devono essere scelti in modo da rendere accettabile tale perdita (teorema del campionamento di Shannon)



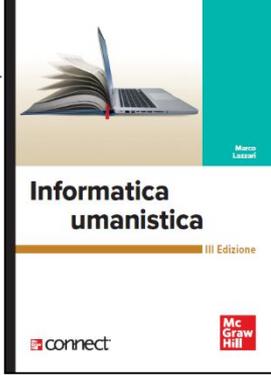
Codifica binaria di altri tipi di informazione

- caratteri: ASCII, ASCII esteso, EBCDIC, UNICODE
- immagini: rappresentazione di un'immagine mediante pixel (bitmap)
 - profondità di colore (color depth): numero di bit per codificare il colore di ogni pixel
- filmati:
 - es. CD: stereofonia, quantizzazione: 65.536 livelli (16 bit), frequenza di campionamento: 44.100 Hz
- filmati: serie di fotogrammi (frame) digitali riprodotti con frequenza costante



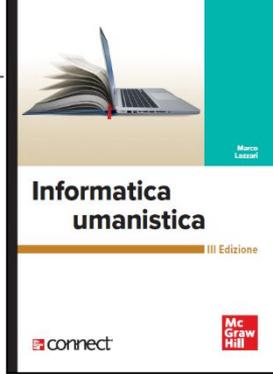
Compressione

- cambiamento di codifica per ridurre il numero di bit utilizzato senza perdere informazione
- compressione lossless: reversibile, adatto a qualsiasi tipo di file, (relativamente) ridotto rapporto di compressione
- compressione lossy: irreversibile, utilizzabile solo in casi specifici, rapporto di compressione variabile a discrezione dell'utente
 - es. immagini JPEG, audio MP3, video MPEG



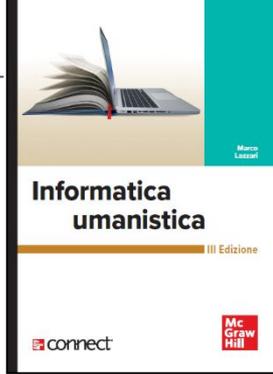
Sistemi informativi e basi di dati

- sistema informativo
- base di dati (DB, Data Base)
- DBMS (Data Base Management System)
 - DDL (Data Definition Language)
 - DML (Data Manipulation Language)
- SQL (Structured Query Language): incorpora un DDL ed un DML, linguaggio per DBMS più diffuso
- modelli per la progettazione di DB
 - modello concettuale
 - modello logico



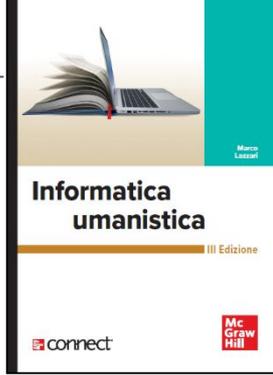
Sicurezza informatica

- cybersecurity
 - riservatezza (privacy)
 - integrità
 - autenticità
 - non ripudiabilità
 - disponibilità
- misure preventive:
 - formazione del personale, protezione dei locali, ridondanza, aggiornamenti software, antivirus, firewall, backup, autenticazione informatica, accountability, comunicazione criptata



Normativa italiana ed europea

- Decreto Legislativo 30 giugno 2003, N. 196 (Codice in Materia di Protezione dei Dati Personali)
 - dato personale
 - dato personale
 - dato sensibile
 - dato giudiziario
 - trattamento
 - informativa e consenso
- Regolamento Generale sulla Protezione dei Dati (RGPD, General Data Protection Regulation, GDPR)



Crittografia

- tecnica utilizzata per la protezione dei dati in trasmissione
 - cifratura: un algoritmo prende un messaggio in chiaro e una chiave di cifratura e genera un messaggio cifrato
 - decifratura: un algoritmo prende il messaggio cifrato e la chiave di decifratura e ricostruisce il messaggio in chiaro
- algoritmi a chiave segreta (simmetrici)
 - chiave di cifratura = chiave di decifratura
- algoritmi a chiave pubblica (asimmetrici)
 - due chiavi: chiave privata (conosciuta solo dal possessore), chiave pubblica (nota a tutti)