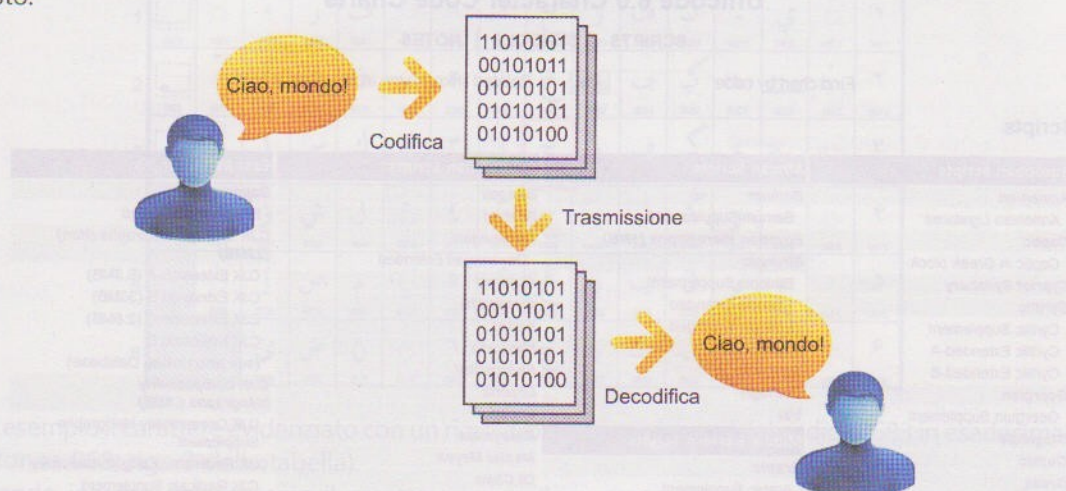


Applicazione

1.7

Utilizzare codici standard per la rappresentazione dei dati

Tutti i dati che si trasmettono da un computer a un altro (o all'interno del computer stesso) sono composti da *byte* (una combinazione di 8 *bit*, cioè di 8 cifre binarie, 0 oppure 1). Quindi ogni trasmissione, che sia un testo, una canzone, un programma o un'immagine, è una lunga sequenza di byte. Per trasmettere un dato, occorre prima convertirlo in binario utilizzando una **codifica**. Il messaggio, dopo essere pervenuto al destinatario, potrà essere decodificato e riconvertito in testo.



È molto importante l'uso di una **codifica standard**, in particolare per i caratteri di testo, in modo tale che tutti gli utenti possano comunicare facilmente e senza errori.

La codifica di base per i caratteri di un testo si chiama **ASCII** (*American Standard Code for Information Interchange*, in italiano *codice americano standard per lo scambio di informazioni*) che utilizza 7 bit per codificare un singolo carattere; per esempio a 0110000 corrisponde il carattere "0" (cifra zero) o a "1110001" corrisponde la lettera "q" (minuscola).

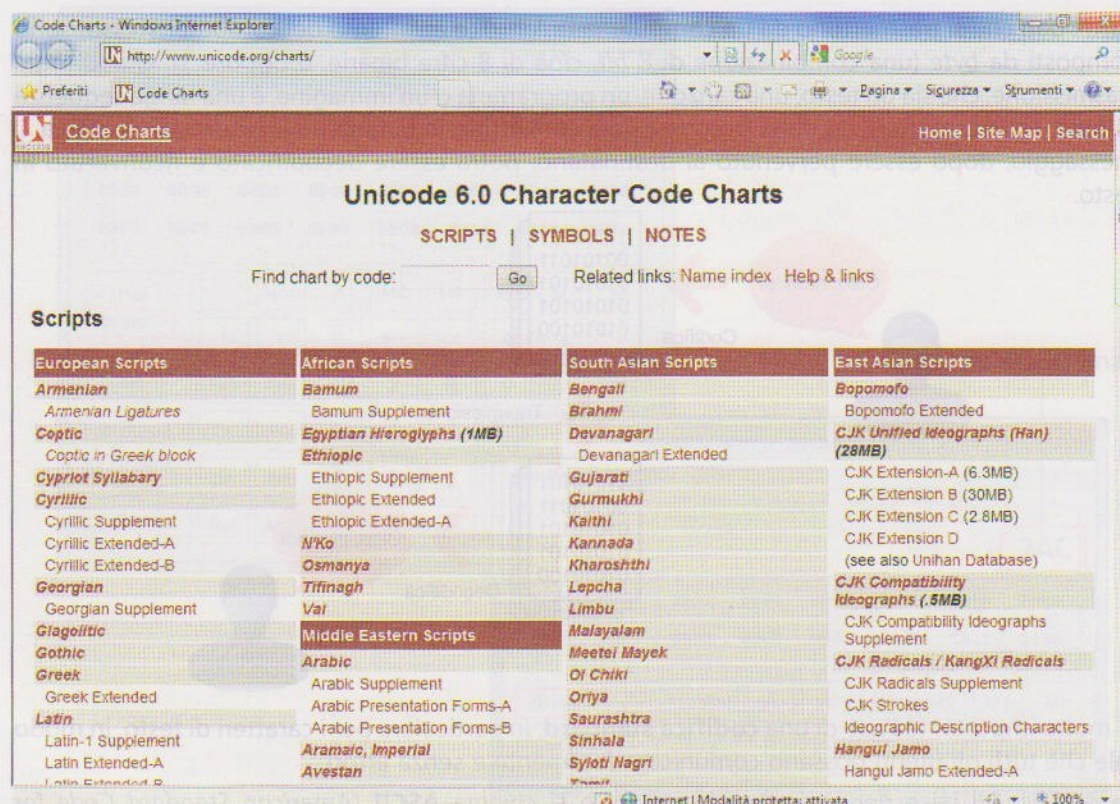
I 7 bit del codice consentono di rappresentare 128 (2^7) caratteri diversi tra loro e quindi non sono sufficienti per codificare tutte le variazioni e tutti i simboli utilizzati nelle varie lingue. Questo sistema è stato esteso a 8 bit, raddoppiando il numero di caratteri disponibili ($256 = 2^8$), con la definizione del codice chiamato **ASCII esteso**.

Nel codice ASCII standard ci sono 32 caratteri (i primi 31 e l'ultimo) **non stampabili**, ma che hanno un significato particolare, per esempio *carattere vuoto*, *fine del testo*, *segnale acustico*, *Cancel*, *Esc*, *Delete*. Successivamente si trovano i caratteri stampabili:

- da **32** (010 0000) a **47**, segni di punteggiatura: punto esclamativo, punto, virgola
- da **48** (011 0000) a **57**, cifre da 0 a 9
- da **58** (010 0000) a **64**, altri segni di punteggiatura: due punti, maggiore, uguale, minore
- da **65** (100 0001) a **90**, lettere maiuscole da "A" a "Z"
- da **91** (101 1011) a **96**, altri segni di punteggiatura: per esempio, apostrofo, parentesi quadre
- da **97** (110 0001) a **122**, lettere minuscole da "a" a "z"
- da **123** (111 1011) a **126**, altri segni di punteggiatura: per esempio, parentesi graffe, tilde.

Nel 1991 è stata sviluppata una nuova codifica, chiamata **Unicode**, che si è diffusa rapidamente ed è in continua evoluzione: è diventata lo standard di fatto per la rappresentazione delle informazioni nei documenti elettronici, in particolare nelle pagine Web, utilizzando i simboli delle numerose lingue esistenti nel mondo.

Le tabelle dei codici *Unicode* sono disponibili sul sito <http://www.unicode.org/charts>.



I primi caratteri di *Unicode* sono esattamente gli stessi della codifica ASCII, in modo da mantenere la compatibilità con il sistema preesistente. All'inizio la codifica utilizzava 2 byte (16 bit, con la possibilità di codificare 65.536 caratteri), ma poi è stata estesa a 32 bit, permettendo la rappresentazione di più di un milione di caratteri differenti. Ovviamente non sono stati codificati subito tutti i caratteri: essi vengono via via assegnati nelle nuove versioni.

L'obiettivo generale di *Unicode* è di creare una codifica che comprenda tutti i caratteri, con tutte le variazioni possibili, di tutte le lingue esistenti, oltre ai simboli utilizzati in matematica e nelle scienze.

Per semplificare le operazioni sono state poi create versioni ridotte del codice che permettono di scrivere i caratteri di uso più frequente in modo più breve: **UTF-8** (a 8 bit), **UTF-16** (a 16 bit) e **UTF-32** (a 32 bit).

Tutte queste codifiche sono compatibili tra loro almeno per quanto riguarda i caratteri standard (lettere maiuscole, minuscole e numeri), mentre può causare problemi con altri simboli (per esempio le lettere accentate o i simboli di frazione).

Riassumendo, ogni carattere viene convertito in bit per essere trasmesso. Più bit vengono utilizzati, più caratteri differenti possono essere utilizzati e più lunga sarà la sequenza di bit da trasmettere.

Per esempio la stringa di testo "Ciao, mondo!" contenente 12 caratteri (occorre tenere conto di tutti i simboli, compresi lo spazio e il punto esclamativo) occuperebbe 84 bit (12 x 7) se codificata in *ASCII standard*, mentre ne occuperebbe 384 (12 x 32) in *UTF-32*.

La figura seguente mostra la prima parte del documento *Unicode* per la codifica dei caratteri delle lingue arabe (codice *Arabic*): i caratteri corrispondono ai numeri compresi nell'intervallo da 0600 a 06FF in esadecimale.

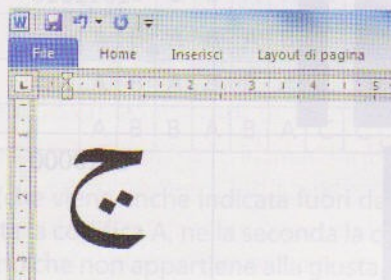
| | 0600 | 0601 | 0602 | 0603 | 0604 | 0605 | 0606 | 0607 | 0608 | 0609 | 060A | 060B | 060C | 060D | 060E | 060F |
|---|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Per esempio il carattere evidenziato con un riquadro rosso corrisponde al codice 0683 in esadecimale (colonna 068, riga 3 della tabella).

Volendo inserire per esempio il carattere in un documento Word, si deve prima convertire in decimale (si può usare la *Calcolatrice* di Windows in *Accessori*):

$$0683_{16} = 1667_{10}$$

Poi, in Word, occorre tenere premuto il tasto **Alt** e premere in successione i tasti 1667 sul tastierino numerico (a destra nella tastiera).



In alternativa, si possono scrivere direttamente nel testo le cifre esadecimali del codice Unicode e premere subito dopo la combinazione di tasti **ALT+X**. Per esempio, scrivendo le cifre 20AC seguite da ALT+X si ottiene il simbolo dell'euro €.

Vediamo ora i codici standard per rappresentare i dati con simboli grafici. Lo scopo di questi codici è di poter acquisire i dati usando strumenti di lettura ottica senza dover digitare cifre o caratteri sulla tastiera. Questo rende le operazioni di input più veloci e più precise.

Il **codice a barre** è una particolare immagine presente sull'etichetta che permette di conoscere le informazioni su un prodotto, per esempio il suo prezzo. Il codice a barre viene acquisito con il **lettore di codice a barre**, come quelli utilizzati nei negozi o alle casse dei supermercati.

Il programma di gestione delle casse interroga, tramite il codice che identifica univocamente un prodotto, il database contenente i dati dei prodotti, ottenendo in risposta la descrizione e il prezzo del prodotto.



Ci sono vari tipi di codici a barre. Il più diffuso è il sistema **EAN** (*European Article Number*) che viene utilizzato per il riconoscimento dei prodotti commerciali.



Il codice è composto da barre verticali nere su sfondo bianco. La barre possono essere di larghezza diverse, così come gli spazi tra di esse. A seconda della diversa larghezza delle barre e degli spazi viene codificata una cifra differente.

La barra nera codifica la cifra 1 mentre lo spazio bianco codifica la cifra 0. Ogni carattere è composto utilizzando 7 posizioni (barre o spazi). Per esempio:

| Figura | Codice binario |
|--------|----------------|
| | 1011010 |
| | 1010000 |

Alcune colonne alle due estremità, destra e sinistra, servono come "inizio", come "fine" e come controllo. Ad ogni combinazione di 7 bit corrisponde una cifra e la codifica dipende anche dalla posizione all'interno del codice. Utilizzare 7 bit per codificare una cifra può sembrare dispendioso (tenuto conto che con 7 bit si può codificare un carattere ASCII) e cambiare la codifica a seconda della posizione può sembrare un'inutile complicazione, ma viene fatto essenzialmente per diminuire gli errori di lettura. La codifica più diffusa, l'**EAN-13** permette di codificare 13 cifre, ma esiste anche l'**EAN-8** che è più semplice e ne codifica solo 8.

Vediamo ora come si compone e come si decodifica il codice a barre con EAN-13.

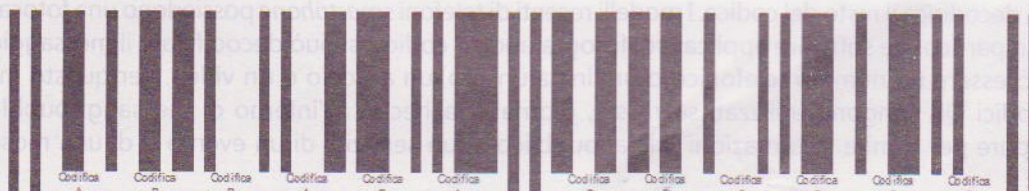
Prima di tutto esistono 3 codifiche, chiamate **A**, **B** e **C**, che utilizzano 7 bit per codificare una cifra:

| Codifica A | | Codifica B | | Codifica C | |
|------------|-------|------------|-------|------------|-------|
| Binario | Cifra | Binario | Cifra | Binario | Cifra |
| 0001101 | 0 | 0100111 | 0 | 1110010 | 0 |
| 0011001 | 1 | 0110011 | 1 | 1100110 | 1 |
| 0010011 | 2 | 0011011 | 2 | 1101100 | 2 |
| 0111101 | 3 | 0100001 | 3 | 1000010 | 3 |
| 0100011 | 4 | 0011101 | 4 | 1011100 | 4 |
| 0110001 | 5 | 0111001 | 5 | 1001110 | 5 |
| 0101111 | 6 | 0000101 | 6 | 1010000 | 6 |
| 0111011 | 7 | 0010001 | 7 | 1000100 | 7 |
| 0110111 | 8 | 0001001 | 8 | 1001000 | 8 |
| 0001011 | 9 | 0010111 | 9 | 1110100 | 9 |

La prima cifra del codice indica quali codifiche utilizzare nelle varie posizioni, come nella seguente tabella.

| Prima cifra | Codifica da utilizzare successivamente | | | | | | | | | | | |
|-------------|--|---|---|---|---|---|---|---|---|---|---|---|
| 0 | A | A | A | A | A | A | C | C | C | C | C | C |
| 1 | A | A | B | A | B | B | C | C | C | C | C | C |
| 2 | A | A | B | B | A | B | C | C | C | C | C | C |
| 3 | A | A | B | B | B | A | C | C | C | C | C | C |
| 4 | A | B | A | A | B | B | C | C | C | C | C | C |
| 5 | A | B | B | A | A | B | C | C | C | C | C | C |
| 6 | A | B | B | B | A | A | C | C | C | C | C | C |
| 7 | A | B | A | B | A | B | C | C | C | C | C | C |
| 8 | A | B | A | B | B | A | C | C | C | C | C | C |
| 9 | A | B | B | A | B | A | C | C | C | C | C | C |

Per esempio se la prima cifra (che viene anche indicata fuori dal codice a barre) è 9, significa che nella posizione 1 viene utilizzata la codifica A, nella seconda la codifica B, nella terza ancora B, ecc. Se viene letto un numero binario che non appartiene alla giusta codifica significa che c'è un errore e occorre rileggere il codice.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|------|---------------|--|---------------|--|---------------|--|---------------|--|---------------|--|---------------|--|---------------|--|---------------|--|---------------|--|---------------|--|---------------|--|---------------|--|--------|--|-----|
|  | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Codifica A | | Codifica B | | Codifica B | | Codifica A | | Codifica B | | Codifica A | | Codifica C | | Codifica C | | Codifica C | | Codifica C | | Codifica C | | Codifica C | | | | |
| | Iniz | Numero | | Numero | | Numero | | Numero | | Numero | | Numero | | Stop | | Numero | | Numero | | Numero | | Numero | | Numero | | Numero | | Fin |
| 9 | = | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | | = | | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | | = |

Dopo la prima cifra si trovano due barre un poco più lunghe (101) che indicano l'inizio del codice. Successivamente si trovano 6 numeri (quindi 42 zone bianche o nere), poi il simbolo di *stop* (01010), altri 6 numeri (42 barre) e il segno di chiusura (101). La codifica, come abbiamo detto, è definita posizione per posizione, a partire dalla prima cifra.

Descriviamo il codice della figura precedente:

- la prima cifra è 9: quindi la codifica da utilizzare per le altre cifre è ABBABA CCCCCC
- poi c'è 101, che è l'inizio del codice (carattere =)
- poi si trova 0011001, che si decodifica, usando la codifica A (come riportato nella tabella vista in precedenza), con **1**
- poi si trova 0011011, che si decodifica, usando la codifica B, con **2**
- seguono altri 4 numeri
- al centro si trova 01010 che viene codificato con =
- seguono altri 4 numeri
- alla fine si trova 101 che viene codificato con = e indica la fine del codice.

Nell'*EAN-13* le prime tre cifre indicano la nazione e le altre codificano univocamente il prodotto.

Il codice *EAN* è un codice a barre *unidimensionale* nel senso che le informazioni sono codificate solamente lungo una direzione. Esistono anche codici a barre *bidimensionali* (codici 2D) formati da piccoli quadretti bianchi e neri e che consentono la codifica di una maggiore quantità di informazioni. Un esempio è il **codice QR** (*Quick Response*, cioè risposta rapida).



La figura è suddivisa in piccoli quadrati bianchi o neri (in modo tale da codificare un bit). Il numero di quadretti è variabile, come la lunghezza del messaggio che è possibile codificare.

La struttura di un codice QR è standard: sono presenti tre quadrati più grossi negli angoli superiori e inferiore sinistro e un quadrato più piccolo nell'angolo inferiore destro.

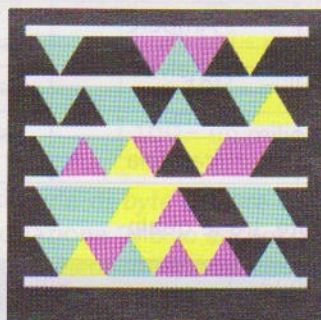
Il lettore di codice QR inizia identificando queste zone particolari per delimitarne le dimensioni e poi decodifica il resto del codice. I modelli recenti di telefoni *smartphone* possiedono una fotocamera e un particolare software applicativo: fotografando il codice, si può decodificare il messaggio che può essere un numero telefonico o un link a un sito, un articolo o un video. Per questo motivo i codici QR vengono utilizzati su riviste, giornali, bacheche all'interno di messaggi pubblicitari oppure per fornire informazioni utili al pubblico di un servizio, di un evento o di una mostra.

Un'altra tipologia di codici 2D è denominata **HCCB** (*High Capacity Color Barcode*, in italiano *codice a colori ad alta capacità*) che si differenzia dal codice QR in quanto utilizza i colori e non solo il bianco e nero. Per esempio, **Microsoft Tag** è un codice che utilizza una griglia 5x10 a quattro colori e consente di memorizzare un link e le informazioni su di esso. Utilizzando i colori aumenta la quantità di dati che è possibile memorizzare. Per esempio un quadratino in bianco e nero può memorizzare un solo bit, associando al bianco la cifra 0 e al nero la cifra 1; lo stesso quadratino in 4 colori può memorizzare 2 bit:

| Colore | Bit |
|--------|-----|
| 1 | 0 0 |
| 2 | 0 1 |
| 3 | 1 0 |
| 4 | 1 1 |

Microsoft fornisce il software di riconoscimento e di codifica per il formato *Tag* per tutte le piattaforme, sia mobili che fisse. Accedendo all'indirizzo <http://tag.microsoft.com> è possibile creare i propri *tag* personalizzati e condividerli con gli altri utenti.

L'applicazione per *smartphone* si può scaricare dal sito <http://gettag.mobi>.



ESERCIZI a pag. 49