Teoria e codifica dell’informazione 12-45

|  |
| --- |
| Informazioni e linguaggio |
| l'informazione è tutto ciò che possiede un significato per l'uomo e che può essere utilizzato immediatamente o conservato in vista di un'utilità futura.  Ľemittente è un soggetto che trasmette l'informazione.  Il ricevente è uno che la riceve e la utilizza.  Per trasmettere un’informazione da un emittente ad un ricevente occorre un supporto fisico. |
| |  |  | | --- | --- | | INFORMAZIONE | SUPPORTO | | direzione da prendere | cartelli stradali | | possibilità di procedere | semaforo | | puntualità dei voli | pagina Web | | fatto accaduto | pagina di cronaca di un quotidiano | | regola grammaticale della lingua inglese | ora di lezione in classe | | determinazione di un'area | formula matematica | |
| Un semaforo con una luce blu lascia perplesso l'automobilista che non è capace di interpretare il segnale, così come una pagina di giornale scritta in cinese o in russo ))) non è comprensibile al comune lettore italiano.  La trasmissione di un'informazione avvenga con un linguaggio che l’emittente conosce e contemporaneamente il ricevente sia in grado di interpretare.  Poi il linguaggio in cui viene espressa l'informazione deve essere convertito in un codice di trasmissione compatibile con il mezzo che si è scelto per la trasmissione stessa.  Se dobbiamo telefonare ad un amico, il linguaggio usato da noi (emittente) e dal nostro amico (ricevente) è quello della lingua italiana, ma, affinché la comunicazione possa avvenire attraverso il cavo del telefono, i suoni emessi dalla nostra voce devono essere convertiti dal nostro apparecchio telefonico in segnali elettrici e dall’apparecchio del nostro amico di nuovo in segnali acustici;  I canale è il mezzo che trasporta l’informazione.  I segnali sono caratteri e i simboli che compongono il messaggio vengono trasformati per viaggiare sul mezzo trasmissivo che collega l'emittente e il ricevente e prendono il nome di segnali. |
| Attraverso un canale di trasmissione, oltre ai segnali che costituiscono l'informazione, vengono anche trasmessi altri segnali che servono a gestire l'informazione: i segnali di "libero" e "occupato" viaggiano sullo stesso canale lungo il quale poi viaggia la nostra voce codificata in segnali elettrici.  Un'informazione deve essere espressa in un linguaggio comprensibile all'emittente e al ricevente e tale linguaggio deve essere tradotto in un codice compatibile con il canale di trasmissione.  Nelle telecomunicazioni i segnali possono essere di tipo digitale quando sono associati a cifre, oppure di tipo analogico quando vengono rappresentati da grandezze che variano con continuità nel tempo. Per esempio, il computer tratta segnali digitali, mentre sulle linee telefoniche viaggiano segnali analogici.  ill problema della comunicazione si complica quando l'emittente è un uomo e il ricevente è una macchina, oppure sia l'emittente che il ricevente sono macchine. In questi casi è necessario che il linguaggio non avrebbe le dubbie interpretazioni.  I linguaggi usati nella trasmissione di informazioni si distinguono in linguaggi naturali e linguaggi formali.  Linguaggi naturali che gli uomini usano per comunicare:   * hanno notevole ricchezza espressiva, * contengono tante ambiguità .   L’uomo ha così costruito dei linguaggi formali:   * basati su simboli e regole ben precise, * privi di eccezioni e di ambiguità, * dedicati di solito a scopi precisi: nell'ambito della matematica, della logica, dell’algebra e della geometria.   Anche se ogni linguaggio ha i suoi termini e le sue regole tuttavia si possono evidenziare alcune caratteristiche comuni.  Ogni linguaggio è costruito su un alfabeto di simboli convenzionali условные detti caratteri.  Il linguaggio della logica usa le lettere dell'alfabeto per indicare le proposizioni, i simboli -i, a, v, <-» per indicare i connettivi.  Una stringa o parola è una successione di caratteri dell’alfabeto.  Il numero 724 è una stringa dell'alfabeto dell'aritmetica; l’espressione a v b è una stringa dell'alfabeto della logica.  Vi sono alcune stringhe che hanno un significato nel linguaggio considerato e non lo hanno se sono scritti in un mdo diverso. Nell'alfabeto dell’aritmetica:   * la stringa x 34 non ha significato * la stringa 3x4 ha significato   la semantica di quel linguaggio è l'insieme dei significati che devono essere attribuiti alle stringhe.  Per costruire un insieme di stringhe, o frase, che abbia significato devono essere stabilite le regole che costituiscono la sintassi del linguaggio.  Il lessico di un linguaggio permette di stabilire se una stringa appartiene o meno al linguaggio considerato.  La grammatica è l'insieme delle regole che consentono di generare le parole e le frasi di quel linguaggio.  I dati sono le conoscenze elementari che caratterizzano una situazione reale e che possono essere conservate per un futuro utilizzo.  I dati sono quindi gli elementi costitutivi delle informazioni.  Ľ elaborazione dei dati il compiuter esegue per ottenere le informazioni.  L'esempio più semplice di elaborazione dei dati è la loro semplice trascrizione (lettura del saldo e comunicazione); un’elaborazione più complessa è quella che prevede di compiere su di essi operazioni di tipo matematico (somme e sottrazioni, confronti e così via).  Ogni elaborazione prevede:   * dei dati di ingresso (dati di input) * produce dei dati in uscita (dati di output). |
|  |
| Nel caso del esempio la carta Bancomat : il codice segreto sono i dati di input, il saldo è un dato di output.  Il risultato del lavoro di elaborazione è costituito da altri dati, che possono essere a loro volta utilizzati per altre elaborazioni a un livello superiore.  Inoltre possiamo riconoscere una distinzione tra il dato e l’informazione nel senso che:   * il dato descrive aspetti elementari di entità o fenomeni; * l'informazione è un insieme di dati elaborati e presentati sulla base dell'esigenza di utilizzo pratico da parte delle persone interessate. |

Comunicazione e interfaccia utente

La teoria della comunicazione :

* si occupa della trasmissione di segnali,
* comprende lo studio della comunicazione tra sistemi della stessa natura: comunicazione tra esseri umani o tra sistemi informatici,
* comprende lo studio della comunicazione tra sistemi della natura differente: come un sistema informatico e un essere umano possono comunicare tra loro.

Gli elementi fondamentali della teoria della comunicazione sono:

* il mittente, che invia il messaggio:
* il destinatario (o ricevente), che riceve il messaggio;
* il messaggio che viene inviato.

La presenza di questi tre elementi non garantisce la corretta trasmissione. Affinché la comunicazione avvenga correttamente, devono essere ben definiti:

* un contesto, cioè tutto ciò che fa da contorno alla comunicazione (luogo, circostanze, eventi, ecc);
* un codice comune a destinatario e mittente (due esseri umani devono saper parlare una lingua comune);
* un contatto, o un mezzo trasmissivo, cioè una connessione (fisica, verbale, psicologica, ecc.) che permetta di stabilire una comunicazione.

Utilizzando i programmi con interfaccia grafica la comunicazione avviene tra un essere umano e un sistema informatico.

Un programma può comunicare in molti modi, per esempio visualizzando un testo oppure emettendo un suono.

Le interfacce dei software si suddividono in due categorie:

* interfaccia testuale, quando la comunicazione avviene esclusivamente via testo. In questo caso anche l’utente può comunicare esclusivamente utilizzando una tastiera.
* interfaccia grafica (GUI, Craphical User Interface), nei casi in cui i programmi utilizzano finestre, icone, mouse e puntatori.

I programmi con interfaccia grafica possono coinvolgere più sensi: un semplice videogioco utilizza sia immagini che suoni, mentre altri sistemi sono sensibili al tocco di un dito.

Quando si crea un programma dotato di interfaccia grafica occorre verificare:

* che funziona correttamente,
* che la comunicazione sia semplice e immediata.

Il programma viene sottoposto a varie tipologie di test che verifichino il corretto funzionamento dell'interfaccia:

* Esistono vari sistemi automatici di controllo (CUI testing tools) che esaminano possibili comportamenti dell’utente e delle risposte del programma.

3. Sistemi di numerazione posizionali

Le quantità numeriche vengono espresse generalmente utilizzando il sistema di numerazione decimale, che si chiama così perché utilizza 10 cifre (0,1,2,3,4,5.6.7.8.9) per rappresentare i numeri. Si dice anche che i numeri sono rappresentati in base 10.

Il sistema di numerazione è un insieme di oggetti e di regole organizzati per svolgere la funzione costituita dalla rappresentazione delle grandezze numeriche.

Il sistema risulta definito:

* dalle parti (le cifre),
* dalle relazioni (le regole di calcolo)
* dalle finalità (rappresentare le grandezze).

Le cifre possiedono un valore posizionale, cioè un valore a seconda della posizione occupata nella scrittura del numero.

Per esempio, nel numero 724 la cifra 7 vale 700, 7 centinaia (potenza di 10 con esponente 2), la cifra 2 vale 20, 2 decine (potenza di 10 con esponente 1) e la cifra 4 vale 4 unità (potenza di 10 con esponente 0).



In informatica vengono usati altri sistemi di numerazione:

* il sistema binario (numeri in base 2),
* il sistema ottale (numeri in base 8)
* il sistema esadecimale (numeri in base 16).

In particolare il sistema binario si lega strettamente alla tecnologia del funzionamento dei computer (funziona usando due cifre 0 e 1).

Nel linguaggio informatico le cifre binarie vengono indicate con il termine bit, dalla contrazione delle due parole inglesi Binary digiT (cifra binaria).

Poiché i dati all'interno del computer sono rappresentati utilizzando le cifre binarie, il bit diventa l’unità elementare, la più piccola unità di informazione.

In modo analogo al sistema decimale, le cifre del sistema binario 0 e 1 assumono un valore posizionale nella scrittura del numero binario con riferimento alle potenze di 2, anziché alle potenze di 10.

Le potenze di 2 con esponente da 0 a 10 sono:



Quindi il numero binario 1001 corrisponde al numero decimale 9, perché



La regola della trasformazione di un numero binario in un numero decimale:

* si moltiplica ciascuna cifra binaria per la corrispondente potenza di 2 e si sommano i prodotti ottenuti.

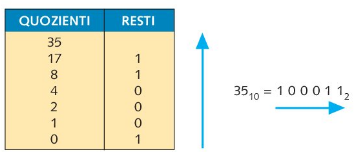
Per esempio, dato il numero binario 110 10:



La regola della trasformazione inversa, da un numero decimale a un numero binario:

* Si divide il numero dato per 2 e si scrive il resto (che può essere 0 o 1);
* il quoziente ottenuto viene a sua volta diviso per 2 ottenendo un nuovo resto;
* si prosegue fino a quando si ottiene come quoziente il numero 0.

La sequenza dei resti è letta dall'ultimo al primo.



Sistema ottale

Il sistema ottale si chiama così perché utilizza 8 cifre nella rappresentazione dei numeri: 0,1,2,3,4,5,6,7. Il valore posizionale delle cifre è fornito dalle potenze di 8:



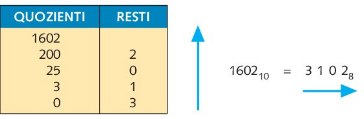
Le regole per la conversione di un numero ottale in decimale e da decimale ad ottale sono analoghe a quelle già descritte per il sistema binario.

La trasformazione del numero ottale 325 in numero decimale:





La trasformazione del numero decimale 1602 in numero ottale:



Si osservi che ogni cifra del sistema ottale può essere rappresentata con una tripletta di bit come mostrato nella seguente tabella.



Da binario a ottale

Si raggruppano le cifre del numero binario a gruppi di tre a partire da destra, e si trasformano le cifre di ciascun gruppo nel corrispondente numero ottale.

Il numero binario 1011110111 si può scrivere come





Da ottale a binario

Ciascuna delle cifre ottali compone il numero un gruppo di tre bit. Il numero ottale 625

si può scrivere 

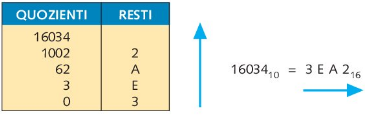
quindi: 6258 = 1100101012



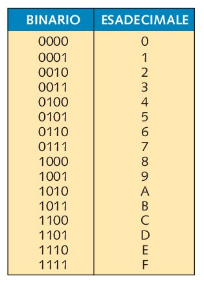
Nn numero da esadecimale a decimale e viceversa: 3AF2



Un numero decimale 16034 in numero esadecimale:



Un numero dal sistema in base 2 al sistema in base 16:



Da binario a esadecimale

Si raggruppano le cifre del numero binario a gruppi di quattro a partire da destra, e si trasformano le cifre di ciascun gruppo nel corrispondente numero esadecimale, secondo la tabella di conversione. Un numero binario 1011110111

si può scrivere come



quindi: 10111101112 = 2F716

Da esadecimale a binario

Si fa corrispondere a ciascuna delle cifre esadecimali che compongono il numero un gruppo di quattro bit secondo la tabella di conversione.

Per esempio, dato il numero esadecimale C3B

si può scrivere



quindi: C3B16 = 1100001110112

4 Rappresentazione dei numeri

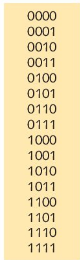
Le cifre binarie 0 e 1 (bit) sono gli elementi fondamentali della scienza dei calcolatori per la rappresentazione delle informazioni all'Interno dell'elaboratore.

Per essere elaborate, le cifre binarie vengono raggruppate in sequenze o stringhe di 8 bit, cioè un byte.

La parola è una stringa di un numero fisso di byte per trattare i dati dagli elaboratori.

Esistono elaboratori che operano con parole di: 1 byte ( 8 bit), 2 byte (16 bit), 4 byte (32 bit), 8 byte (64 bit).

Il tempo per elaborare una parola va considerato come costante e quindi la lunghezza delle parole costituisce uno degli indici per determinare la potenza e la velocità di calcolo degli elaboratori.

Dati N simboli da disporre in un numero fisso P di posti, tutte le possibili combinazioni sono in numero pari a Np.

Utilizzando i due simboli delle cifre binarie, le combinazioni che si ottengono con 4 posti sono:

16 in totale (con 2 simboli in 4 posti si ottengono 24 combinazioni diverse).

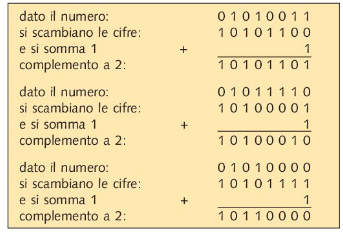
Utilizzando i 16 simboli del sistema esadecimale da disporre in 2 posti, si ottengono 162 = 256 combinazioni.

Rappresentazione dei numeri interi

Supponiamo di utilizzare parole di 2 byte (16 bit) e prendiamo in considerazione soltanto numeri interi con segno. Dei 16 bit a disposizione il primo bit a sinistra viene dedicato alla rappresentazione del segno: 0 per il segno +, 1 per il segno -.

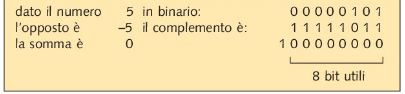
Per i numeri positivi si hanno, quindi, a disposizione 15 bit, cioè 2 simboli (le cifre binarie), da disporre in 15 posti: possiamo quindi rappresentare 215 numeri diversi, vale a dire dal numero 0 al numero 215 - 1 = 32767.

Per i numeri negativi, invece, si ricorre al criterio del complemento a 2: dato un numero espresso in forma binaria, per ottenere il complemento a 2 del numero dato, ad ogni cifra si sostituisce 0 dove c'è 1, 1 dove c'è 0 e alla fine si somma 1.



Sussiste la seguente regola: fissato il numero di bit su cui operare, e convenendo di ignorare i riporti che creino nuove cifre a sinistra eccedenti il numero di bit fissato, il complemento a 2 di un numero dato risulta l'opposto del numero.

Per esempio, fissato a 8 il numero di bit su cui operare:



Il riporto a sinistra degli otto bit utilizzati viene ignorato.

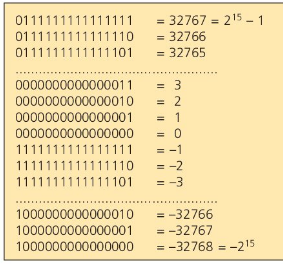
Per calcolare l'opposto di un numero binario si utilizza il complemento a 2 del numero. Utilizzando quindi parole di 2 byte e volendo rappresentare i numeri interi positivi, poiché il primo bit di sinistra è dedicato all'indicazione del segno, si ottiene che i numeri interi positivi rappresentabili in 2 byte vanno da 0 a 215 - 1.

Per i numeri negativi, invece, si utilizza il criterio del complemento a 2 e quindi i numeri negativi rappresentabili in 2 byte vanno da -1 a -215 .

Infatti, utilizzando 16 bit:

il numero 1 viene rappresentato da 0000000000000001 объяснения смотреть тут http://infolike.narod.ru/codir2.html

il numero -1 dal complemento di 1, cioè 1111111111111111 il numero +215 in binario risulta uguale a 1000000000000000 e non può essere rappresentato in una parola di 2 byte perché il primo bit a sinistra indica il segno Invece il complemento a 2 del numero precedente, ovvero l’opposto di 215, risulta ancora 1000000000000000 = -215. Riassumendo, i numeri interi rappresentabili in una parola di due byte sono (in ordine decrescente)



Il primo e l'ultimo numero della tabella precedente sono

215 - 1; - 215.

L'esponente di 2 è 15 perché si considerano parole di 16 bit di cui il primo è utilizzato per il segno. Nel caso si utilizzino parole di 4 byte, i bit a disposizione risultano 32 di cui uno è dedicato al segno e quindi:

massimo = 231 - 1 e minimo = - 231

5. Rappresentazione delle Informazioni alfanumeriche

Con un bit si hanno due combinazioni possibili: 0, 1. Con un byte abbiamo 256 combinazioni possibili: da 00000000 a 11111111.

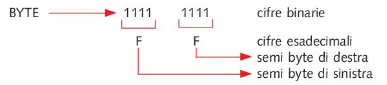
Infatti 11111111 in base 2 = 128 + 64 + 32 + 16 + 8 + 4 + 2 + 1= 255.

Il massimo valore che può esprimere un byte è 255 decimale.

Il massimo numero di combinazioni che può contenere un byte è 256 (compresa la combinazione 00000000).

Un byte nel quale tutti i bit sono uguali a 1 (11111111) è equivalente a FF esadecimale.

Per rappresentare un simbolo del sistema esadecimale occorrono 4 bit (4 cifre binarie).



La rappresentazione di una cifra esadecimale richiede un semibyte.

Tutte le informazioni non numeriche (alfabetiche o alfanumeriche) sono esprimibili mediante una combinazione di lettere, cifre o caratteri speciali: affinché un elaboratore riesca a riconoscere e a trattare tali informazioni deve essere stabilita una corrispondenza che ad ogni carattere utilizzato per rappresentare le informazioni associ una particolare configurazione degli 8 bit di un byte.

La codifica è l’associazione di una combinazione di bit ad un determinato simbolo (lettera, cifra o carattere speciale)

La codifica 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.

I 7 bit del codice consentono di rappresentare 128 (27) 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 = 28), 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, Canee!, Esc, Delete. Successivamente si trovano i caratteri stampabili:

* da 32 (010 0000) a 47, il carattere spazio e segni di punteggiatura quali: 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: per esempio, 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.

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).

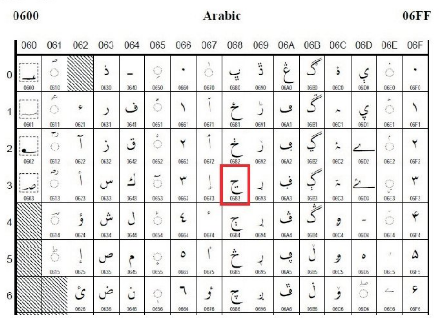
Riassumendo:

* 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 Arabie): i caratteri corrispondono ai numeri compresi nell'intervallo da 0600 a 06FF in esadecimale.



Per esempio :

* il carattere evidenziato con un riquadro rosso corrisponde al codice 0683 in esadecimale
* 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): 068316 = 166710
* Poi, in Word, premere il tasto Alt e premere 1667 sul tastierino numerico (a destra nella tastiera).

In alternativa:

* si possono scrivere direttamente nel testo le cifre esadecimali del codice Unicode subito dopo la combinazione di tasti ALT+X (ALT+X + 20AC si ottiene il simbolo dell'euro €.)

6.Formato delle immagini

La carta stampata, la televisione, la radio e il Web utilizzano canali diversi e quindi formati diversi per i contenuti.

Un’ immagine per essere stampata

* ecessita di una quantità di particolari, di colori e di informazioni molto maggiore rispetto ad un’immagine destinata ad un sito Web,

Un' immagine destinata ad unsito Web:

* dove l'importante una buona qualità con un caricamento il più rapido possibile, riducendo le dimensioni dell'immagine.

Lo stesso vale per i video che, nel caso del Web, devono essere ottimizzati e ridimensionati per assicurare uno streaming o un download veloce. Diversamente, un filmato creato per la televisione dovrà assicurare una qualità maggiore, vista anche la disponibilità di televisori full HD (High Definition).

I formati grafici sono solitamente classificati secondo due tipologie:

* grafica vettoriale
* grafica raster.

Le immagini vettoriali sono descritte da un insieme di primitive geometriche (punti, linee, segmenti, poligoni, curve) alle quali è possibile assodare colori di riempimento, colori dei contorni e sfumature.

Le proprietà di un'immagine, quali il contorno e il colore di riempimento, sono elementi indipendenti che possono essere spostati e modificati usando un software .

Le immagini vettoriali sono più leggere delle immagini raster perché per la loro rappresentazione sono necessarie poche informazioni che possono anche essere compresse.

Osservando un’immagine raster ed una vettoriale rappresentanti lo stesso soggetto, la versione vettoriale (figura a sinistra) denota una qualità maggiore rispetto alla versione raster (figura a destra).

I programmi per la creazione e la modifica di immagini vettoriali sono:

* Adobe lllustrator, Corel Draw (prodotti commerciali) e il software libero Inkscape (open source).

Il formato più diffuso per le immagini vettoriali è .svg (Scalable Vector Graphics):

* È il formato standard del W3 Consortium
* è visualizzabile dai browser solo attraverso l'installazione di plug-in aggiuntivi.

Le versioni recenti dei browser sono in grado di interpretare il codice per implementare la grafica vettoriale e quindi visualizzano correttamente questo tipo di immagini.

Un altro formato vettoriale molto usato è eps (Encapsulated PostScript):

* è detto ibrido, perché può contenere sia immagini vettoriali che non vettoriali.

Altri formati vettoriali sono proprietari, cioè legati a software specifici:

* .swf (Shockv/ave Flash) relativo alle animazioni create con il software Adobe Flash, .
* ai per il software Adobe lllustrator
* cdr per il software Corel Draw.

Tra i formati vettoriali ibridi occorre ricordare anche il formato .pdf (Portable Document Formai), molto utilizzato più per i documenti che per le sole immagini. I file .pdf vengono creati e modificati con il software Adobe Acrobat.

Con la diffusione del Web, era necessario poter disporre di un formato che potesse gestire immagini, testi e collegamenti ipertestuali, ma che non potesse essere modificato. I file creati con i programmi di word processing potevano contenere tutti questi elementi, ma restavano modificabili.

Uno scambio documentale su vasta scala richiede la trasmissione di informazioni in modo da preservare la fonte e gli originali. Ma:

* permette funzioni avanzate di ricerca, selezione, stampa e zoom, la possibilità di inserire commenti e revisioni e firme digitali. Tutto questo lasciando inalterato il documento originale.

Per libri con molto testo si utilizza il formato ePub (electronicpublishing, pubblicazione elettronica):

* è uno standard per la produzione di e-book (electronic hook, libro elettronico o libro digitale)
* funziona con qualsiasi dispositivo mobile di lettura, e-reader, tablet o smartphone.

Le immagini raster (dette anche bitmap) sono formate da singoli punti, i pixel (picture element). Ogni pixel che compone un'immagine assume un colore; l’insieme dei pixel colorati genera l'immagine.

La grafica raster si usa per le immagini fotografiche, perché riesce a gestire un elevato numero di colori. Maggiore è il numero di colori, maggiore sarà il peso in Kilobyte e Megabyte dell’immagine, e quindi la dimensione del file.

Aumentando le dimensioni di un'immagine raster la qualità peggiora notevolmente: il software grafico deve aggiungere pixel nuovi accanto a quelli già esistenti, individuando la tonalità giusta con una sfumatura simile a quella dei pixel vicini, attraverso un procedimento chiamato interpolazione.

Per questo motivo le immagini raster sono dipendenti dalla risoluzione, a differenza delle immagini vettoriali che sono indipendenti dalla risoluzione e quindi scalabili.

Tra i software più noti per la modifica delle immagini raster :

* Adobe Photoshop e The Gimp (il secondo è un software libero).

I formati più comuni delle immagini raster sono: .

* jpg (Joint Photographic Experl Group), .
* gif (Graphic Interchange Format),
* .png (Portable Network Graphics)
* .bmp (Windows bitmap).
* .tif
* .raw

In ambito professionale più usati i formati .tif e .raw che consentono di operare modifiche senza perdita di qualità, anche se la dimensione delle immagini è maggiore rispetto agli altri formati.

|  |  |
| --- | --- |
| Confronto tra raster e vettoriale | |
| Grafica raster | Grafica vettoriale |
| formata da pixel | formata da primitive geometriche |
| grafica elaborata, fotografie a milioni di colori | grafica semplice, ideale per disegni e loghi |
| modificando le dimensioni la qualità peggiora: è dipendente dalla risoluzione | modificando le dimensioni la qualità resta inalterata: è scalabile |
| pieno supporto di tutti i browser | necessita di plug-in |
| a parità di soggetto la dimensione dell'imma- gine è maggiore, ma visualizza più colori | a parità di soggetto la dimensione dell'immagine è minore, ma la resa cromatica è più limitata |

I formati per le immagini Web

I browser sono in grado di visualizzare correttamente i tre formati grafici raster:

* .jpg, .gif .png.

Tuttavia le loro caratteristiche sono diverse.

formato .gif:

* può rappresentare al massimo 256 colori (utilizza 8 bit),
* è adatto alla visualizzazione di loghi e disegni con pochi colori e senza sfumature.
* La sua caratteristica più apprezzata, è la apacità di gestire le trasparenze: l'immagine con lo sfondo trasparente.
* Un'altra caratteristica del formato .gif è la possibilità di salvare un insieme di immagini in successione creando un'animazione (gif animata).
* utilizzano una compressione lossless (senza perdita di qualità)

formato .jpg

* è adatto alla rappresentazione di immagini fotografiche: si basa sul modello RGB quindi può rappresentare immagini con sfumature di milioni di colori (utilizza 24 bit).
* Non supporta né le trasparenze né le animazioni.
* utilizzano una compressione lossy (con perdita di qualità)

formato .png:

* unisce i vantaggi della gestione di milioni di colori del formato .jpg con la possibilità di impostare le trasparenze del formato .gif.
* utilizzano una compressione lossless (senza perdita di qualità)

Formati audio e video

Perché il suono venga rappresentato in un file è necessario che venga digitalizzato, trasformando il segnale da analogico a digitale (processo di trasformazione è detto campionamento).

La tipologia di compressione:

I formati audio si dividono in formati compressi (lossy) e non compressi (lossless).

* I formati compressi occupano meno spazio di quelli non compressi, ma la qualità dell'audio è peggiore.

Alla categoria lossless appartengono i formati .wav e .aiff, generalmente di grandi dimensioni, ma di alta qualità. Sono molto diffusi nell'ambito della registrazione audio:

II formato .wav (contrazione di WAVeform audio file format) :

* è lo standard di Windows per i file sonori.

Il formato .aiff (Audio Interchange File Format)

* è analogo al .wav. ma è sviluppato da Apple.

I formati .mp3 (Mpeg 1 Layer 3) e .wma (Windows Media Audio), grazie alla compressione, permettono di ottenere file audio di dimensioni molto minori rispetto ai formati non compressi, con una riduzione della dimensione che può arrivare anche a 1/12 rispetto a un file .wav.

Il formato .wma

* è il formato predefinito per le registrazioni effettuate tramite il registratore di suoni di Windows 7 e 8.

Il formato .mp3

* taglia dal file audio le informazioni ritenute non necessarie: poiché l'orecchio umano viene ritenuto non in grado di percepire determinate frequenze, le frequenze alte vengono eliminate.
* La qualità di ascolto, è inferiore rispetto ad un CD audio, è discreta in bitrate (numero di bit utilizzati per memorizzare un secondo di audio) viene mantenuto al di sopra dei 128 Kb/s (kilobit al secondo).

Un formato Real Audio:

* È adatto all'ascolto di file in streaming con il buon livello di compressione.

Il formato midi (Musical Instrument Digital Interface) .mid

* è diverso dagli altri perché non contiene suoni, ma esclusivamente le istruzioni per suonare le note, con i relativi tempi e gli strumenti musicali che devono eseguirle, proprio come su uno spartito.
* I file con estensione .mid sono modificabili nota per nota
* i file midi hanno una dimensione molto ridotta.
* Il formato midi viene utilizzato per la produzione di musica elettronica e di basi musicali.

I filmati sono costituiti da fotogrammi (trame) messi in rapida sequenza.

I televisori e i monitor possiedono un rapporto di larghezza/altezza(aspect ratio) dell'immagine di 4:3 oppure di 16:9. Inoltre ogni immagine che viene trasmessa è formata da un numero di fotogrammi al secondo (l’unità di misura è fps, frame per second) che varia a seconda dell'area in cui ci si trova:

* In America del Nord e in Giappone lo standard è di 30 fps e il sistema è chiamato NTSC.
* In Europa lo standard è chiamato PAL e si basa su 25 fps.

Ogni fotogramma non è formato da una sola immagine, ma da due diverse immagini che si sovrappongono completandosi. Questo metodo viene usato per tutti i formati: si dice che il frame è interlacciato.

Con la diffusione dei computer e della rete Internet si è reso necessario, come per l'audio, creare formati che potessero garantire una buona qualità del video, mantenendo gli fps inalterati senza però dover gestire file di grandi dimensioni.

Gli sviluppatori del formato grafico .jpg, hanno creato il primo formato audio/video denominato mpg:

* Questo formato è compatibile con tutti i computer ed è supportato da tutti i sistemi operativi.
* Il formato mpeg-2, usato nei DVD, produce un'altissima qualità, ma necessita di supporti di archiviazione con capacità elevate.
* mpeg-4, è usato per comprimere video di qualsiasi tipo e dimensione, garantendo una qualità alta con una dimensione sensibilmente minore rispetto al formato mpeg-2.

Molto usato in rete è anche il formato .mov (Quicktime):

* realizzato da Apple, ma compatibile anche con i computer Windows.
* garantisce una buona resa grafica e una dimensione ridotta del file.

Il formato .avi (Audio Video Interleave):

* è stato realizzato da Microsoft nel 1992 come formato standard per i sistemi operativi Windows.
* Permette un'ottima qualità al pari del successivo formato .wmv (Windows Media Video), ideato sempre da Microsoft,
* adatto anche allo streaming video.

Il sistema codec (acronimo di compressore-decompressore):

* Tutti questi formati devono utilizzare un sistema per la compressione dei video senza ridurne le dimensioni.
* Un codec è un insieme di istruzioni che consentono di comprimere e decomprimere un file video o audio.
* Anche per acquisire un video digitale viene effettuata una campionatura: per i video non si tratta di memorizzare solo le onde sonore, nel caso sia presente l'audio, ma anche il colore dei pixel.
* Con una codifica a 8 bit si visualizzano solo 256 colori, mentre usando 24 bit. con il sistema RGB, i colori disponibili sono 16 milioni, ottenendo un risultato simile all’originale analogico.
* Anche per il video la compressione può essere con perdita (lossy) e senza perdita (lossless).

l codec più utilizzati sono Divx e Xvid utilizzati dal formato .avi e caratterizzati da una buona qualità e dimensioni contenute. Essi consentono anche la scelta del livello di compressione:

* mpeg-1 (video CD),
* mpeg-2 (DVD)
* mpeg-4 per i file .mpg.

Inoltre il codec H264 (evoluzione di mpeg-4) offre un'alta qualità, simile a quella del DVD, ma con una dimensione notevolmente inferiore. Questo codec è attualmente usato per i due formati dei DVD ad alta definizione, HD-DVD e Blu-ray.

Codifica a barre e QR

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.

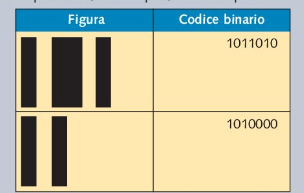
Il programma di gestione delle casse interroga, tramite il codice che identifica univocamente un prodotto.

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

0 0123451678905

Il codice:

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



Alcune colonne alle due estremità, destra e sinistra, servono :

* come “inizio",
* come “fine"
* come controllo.

Ad ogni combinazione di 7 bit corrisponde una cifra e la codifica dipende anche dalla posizione all'interno del codice. La codifica più diffusa, l'EAN-13 permette di codificare 13 cifre, ma esiste anche I’EAN-8 che è più semplice e ne codifica solo 8.

Il codice a barre con EAN-13:

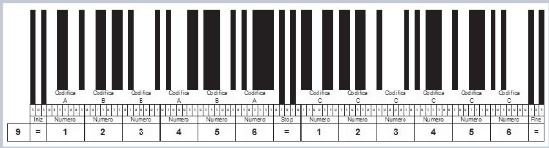
Prima di tutto esistono 3 codifiche, chiamate A, B e C, che utilizzano 7 bit per codificare una cifra:



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



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.



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.

Nella’ 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, 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.

La codifica QR è classificata in diverse versioni, dalla versione 1 alla versione 40. Ogni versione ha una differente configurazione di moduli e un differente numero di moduli.

I moduli di QR sono il modulo che si riferisce ai punti bianchi e neri che compongono il codice QR.

Il termine configurazione dei moduli si riferisce al numero di moduli contenuti in un codice QR, dalla versione 1 che contiene 21 x 21 moduli, alla versione 2 con 25 x 25 moduli, fino alla versione 40 con 177 x 177 moduli.

Ogni versione di livello superiore aggiunge 4 moduli per lato.

Per ogni versione si può stabilire una capacità massima di dati rappresentabili.

I codici QR possono memorizzare fino a un massimo di 4296 caratteri alfanumerici oppure 7089 caratteri numerici.

Un sistema QR Code è formato da 2 parti:

* da un lato il software che genera il codice QR e il dispositivo che lo stampa;
* dall'altro lato lo scanner (o la fotocamera di un dispositivo mobile) e l'app che legge il codice decodificandolo.