Introduzione alla codifica dell'informazione

Contenuto

- Codici e Codifiche
- Codici numerici
- Rappresentazione dei numeri naturali
- Rappresentazione dei numeri relativi
- Aritmetica binaria, somma e sottrazione
- La costruzione di una ALU e sue ottimizzazioni
- Aritmetica binaria, moltiplicazione e divisione
- Circuiteria a supporto di moltiplicazione e divisione
- Rappresentazione dei numeri razionali
- Aritmetica binaria su numeri razionali

Introduzione

- La codifica dell'informazione ha origine intorno agli anni '50 grazie a Turing, Von Newmann, Shannon.

- Oggi è strettamente legata all'elaborazione automatica
  - Codifica numerica
  - Crittografia
  - Compressione
  - Rilevamento e correzione di errori

- In questo corso si discutono i fondamenti della codifica di informazioni di tipo numerico finalizzata all'elaborazione automatica.

Codici e Codifiche

- Un codice è un elemento sintattico che può essere descritto come insieme delle parole che lo costituiscono:

$$C = \{ X_1, X_2, \dots , X_n \}$$

- Ogni parola X è costituita da una sequenza di simboli a_j e l'insieme di tali simboli costituisce l'alfabeto A del codice

$$A = \{ a_1, a_1, \dots , a_k \}$$

- ES: numerazione romana:

$$C = \{ I, II, III, IV, \dots , XXIV, \dots \}$$

$$A = \{ I, V, X, L, C, M, D \}$$

Codici e Codifiche - 2

- Per definire un codice serve anche un insieme di regole che permettono di costruire le parole
  - Ad es IIV e MXXC non appartengono alla numerazione romana

- Esistono comunque codici in cui qualsiasi combinazione di simboli dell'alfabeto è ammissibile
  - Es: numerazione arabo

$$A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

qualsiasi combinazione di questi simboli produce una parola valida chiamata numero

codifichiamo i giorni della settimana

- I codici sono utilizzati per rappresentare un insieme di informazioni
  - Numeri, colori, mesi dell'anno, …
- Ogni relazione che associa gli elementi dell'insieme delle informazioni alle parole del codice prende il nome di codifica
- ES: codifichiamo i giorni della settimana
  - Usando un alfabeto composto da tre simboli è possibile costruire nove parole di lunghezza due
  - Delle quali ne useremo in maniera arbitraria solo sette

codifichiamo i giorni della settimana -2

I = { lunedì, martedì, mercoledì, giovedì, venerdì, sabato, domenica }

A = {a,b,c}

C={aa,ab,ac,ba,bb,bc,ca,cb,cc}

- Definire una codifica significa stabilire una associazione tra l'insieme *C* e l'insieme *I*
- Queste sono due possibili codifiche:

lunedì	→	*aa*
martedì	→	*ab*
mercoledì	→	*ac*
giovedì	→	*ba*
venerdì	→	*bb*
sabato	→	*bc*
domenica	→	*ca*

lunedì	→	*aa*
martedì	→	*bb*
mercoledì	→	*ba*
giovedì	→	*bc*
venerdì	→	*ac*
sabato	→	*cb*
domenica	→	*ca*

Codifica univoca o ambigua

- Una codifica può essere non univoca o ambigua
  - ad una stessa informazione sono associate più parole
  - raramente utilizzate nel nostro campo
- Oppure biunivoca
  - ad ogni elemento dell'insieme delle informazioni corrisponde una ed una sola parola del codice
  - es: rappresentazione di informazioni di tipo numerico
    - rappresentazione dei numeri Naturali in base 10
    - notazione scientifica normalizzata (mantissa ed esponente)

Codici e codifiche - classificazioni

- Codici e codifiche possono essere classificati secondo diversi criteri
  - Codici posizionali e non posizionali
    - Nei primi la relazione tra l'informazione e la sua parola di codice può essere espressa tramite una relazione algebrica basata sulla posizione dei simboli all'interno della parola
  - Codici a lunghezza fissa
    - La lunghezza delle parole è una costante prefissata
  - Codici a lunghezza variabile
    - Le parole possono avere lunghezze diverse
    - Molto usati in crittografia
    - Per rappresentare in modo non ambiguo n informazioni con k simboli la lunghezza massima delle parole deve essere $l = \lceil \log_k n \rceil$
    - Es.: 100 informazioni con 3 simboli richiedono $\lceil \log_3 100 \rceil = 5$

      Con 3 simboli e parole di lunghezza 5 si ottengono 3^5=243 sequenze

Codici numerici

- Per rappresentare informazioni di tipo numerico si usano principalmente codici posizionali
- La cardinalità dell'alfabeto $b = |A|$ è detta base della codifica
- Dato $A = \{a_1, a_2, \ldots, a_b\}$ un alfabeto di b simboli una parola di codice posizionale in base b è un sequenza di lunghezza n+k del tipo:
- $X = [\alpha_{n-1} \cdots \alpha_1 \alpha_0, \alpha_{-1} \cdots \alpha_{-k}]$
  - I simboli da α_{n-1} a α_0 rappresentano la parte intera
  - I simboli da α_{-1} a α_{-k} rappresentano la parte frazionaria
- Il valore numerico di una parola X è dato dalla relazione:
- $v(X) = \sum_{i=-k}^{n-1} \alpha_i b^i$

Codici numerici - 2

Codice	Base	Alfabeto
Binario	2	*A*={0,1}
Ottale	8	*A*={0,1,2,3,4,5,6,7}
Decimale	10	*A*={0,1,2,3,4,5,6,7,8,9}
Esadecimale	16	*A*={0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F}

- $v(X) = \sum_{i=-k}^{n-1} \alpha_i b^i$
- Consideriamo come primo esempio la base 10
- $v([38,46]) = 3 \cdot 10^1 + 8 \cdot 10^0 + 4 \cdot 10^{-1} + 6 \cdot 10^{-2} = 30 + 8 + 0,4 + 0,06 = 38,46$
- Una notazione comune è quella di specificare il codice aggiungendo a una parola un pedice che ne indichi la base
  - 11_2è una parola di codice binario mentre 11_{10} è una parola di codice decimale

Rappresentazione dei numeri naturali

- $N = \{1, 2, 3, \dots\}$

- Mediante parole di *n* bit su un alfabeto di *b* simboli è possibile rappresentare b^n valori interi compresi nell'intervallo chiuso
  - $[0; b^n - 1]$

- Poiché i numeri naturali sono privi di parte frazionaria
- $v(X) = \sum_{i=0}^{n-1} \alpha_i b^i$

Conversione di base

- Il procedimento di passaggio tra due rappresentazioni posizionali con basi differenti è detto conversione di base
- Sia
  - $\boldsymbol{A = \{a_0, \dots, a_{b_a}\}}$ l'alfabeto della base $\boldsymbol{b_a}$
  - $\boldsymbol{B = \{b_0, \dots, b_{b_b}\}}$ l'alfabeto della base $\boldsymbol{b_b}$
- Convertire una parola

$$\boldsymbol{X = [\alpha_{n-1} \cdots \alpha_0]_{b_a}}$$

- in un'altra parola

$$\boldsymbol{Y = [\beta_{m-1} \cdots \beta_0]_{b_b}}$$

- significa trovare i coefficienti $\boldsymbol{\beta_j} \in \boldsymbol{B}$ a partire dai coefficienti $\boldsymbol{\alpha_i} \in \boldsymbol{A}$,
- con $\boldsymbol{i = 0..n}$, $\boldsymbol{j = 0..m}$
- **Vediamo due metodi:** somma di prodotti e divisioni ripetute.

Somma di prodotti

- Questo metodo si basa sulla relazione
- $\boldsymbol{valore\ di\ X\ in\ base\ b\ = \sum_{i=0}^{n-1} [\alpha_i]_b \cdot b^i}$ **2.1**
- Il cambiamento dalla base b_a alla base b_b si ottiene esprimendo i coefficienti del vettore
- $X = [\alpha_{n-1} \cdots \alpha_0]_{b_a}$
- in base b_b ed applicando la relazione 2.1 ed effettuando tutte le operazioni nella base $b_b$di riferimento.

Somma di prodotti - esempi

- Esempio 1:
  - $[\,A\,B\,]_{16} \rightarrow [\,?\,]_{10}$
    - Si esprimono i coefficienti A_{16} e B_{16} in base 10
  - $A_{16} = 10_{10} \qquad B_{16} = 11_{10}$
    - Applicando la regola di conversione
  - $[\,A\,B\,]_{16} = 10_{10} \cdot 16_{10}^{1} + 11_{10} \cdot 16_{10}^{0} = [1\,7\,1]_{10}$
- Esempio 2:
  - $[\,A\,B\,]_{16} \rightarrow [\,?\,]_{8}$
  - $A_{16} = 12_{8} \qquad B_{16} = 13_{8}$
  - $[\,A\,B\,]_{16} = 12_{8} \cdot 20_{8}^{1} + 13_{8} \cdot 20_{8}^{0} = [\,2\,5\,3\,]_{8}$
- Difficile operare con basi diverse da 10
- Particolarmente adatto quando la nuova base è 10

Divisioni ripetute

- Questo metodo è un procedimento iterativo che ricava passo dopo passo i singoli coefficienti della parola di codice nella nuova base.
- Sia $\boldsymbol{X} = [\boldsymbol{\alpha}_{n-1} \cdots \boldsymbol{\alpha}_0]_{\boldsymbol{b}_a}$ la parola di codice in base $\boldsymbol{b}_a$ che si vuole convertire nella parola di codice incognita $\boldsymbol{Y} = [\boldsymbol{\beta}_{n-1} \cdots \boldsymbol{\beta}_0]_{\boldsymbol{b}_b}$ in base $\boldsymbol{b}_b$
- Questa volta applichiamo la relazione 2.1 operando però in modo inverso
  - Si parte dal valore di $\boldsymbol{Y}$ in base $\boldsymbol{b}_a$, $\boldsymbol{v}(\boldsymbol{Y}) = \mathbf{X}$
  - si divide per la base $\boldsymbol{b}_b$
  - La base $\boldsymbol{b}_b$ deve essere espressa mediante la base $\boldsymbol{b}_a$
  - le operazioni devono essere svolte in base $\boldsymbol{b}_a$
  - $\boldsymbol{v}(\boldsymbol{Y}) = \boldsymbol{\beta}_{m-1}\boldsymbol{b}_b^{m-1} + \cdots + \boldsymbol{\beta}_1\boldsymbol{b}_b^1 + \boldsymbol{\beta}_0\boldsymbol{b}_b^0$
  - $\frac{\boldsymbol{v}(\boldsymbol{Y})}{\boldsymbol{b}_b} = \boldsymbol{\beta}_{m-1}\boldsymbol{b}_b^{m-2} + \cdots + \boldsymbol{\beta}_1\boldsymbol{b}_b^0 + \frac{\boldsymbol{\beta}_0}{\boldsymbol{b}_b} = \boldsymbol{Q} + \frac{\boldsymbol{R}}{\boldsymbol{b}_b}$
  - In cui Q è il quoziente ed R = $\boldsymbol{\beta}_0$
  - Si ripete il tutto ricavando gli altri coefficienti finché $\mathbf{Q} = \mathbf{0}$

Divisioni ripetute – esempio 1 - qui

- Esempio 2: Convertire $\mathbf{346_{10}}$ in base 8

		8	←Base
$Q^{(0)}$	346	2	←β_0
$Q^{(1)}$	43	3	←β_1
$Q^{(2)}$	5	5	←β_2
$Q^{(3)}$	0		

- Riprova…
- $\boldsymbol{v([5\,3\,2]_8) = 5 \cdot 8^2 + 3 \cdot 8^1 + 2 \cdot 8^0 = 320 + 24 + 2 = 346_{10}}$

Divisioni ripetute – esempio 2

- Esempio 1:
  - $[\,2\,5\,3\,]_8 \rightarrow [\,?\,]_{16}$
    - Si deve eseguire la divisione ripetuta per $\boldsymbol{b_b = 20_8}$
  - $\mathbf{253_8 : 20_8 = 12_8 + \frac{13_8}{20_8}}$, $\quad \mathbf{12_8 : 20_8 = 0_8 + \frac{12_8}{20_8}}$
    - Si convertono i resti $\mathbf{13_8}$ e $\mathbf{12_8}$ nella nuova base 16 ottenendo
  - $\mathbf{13_8 = \boldsymbol{B}_{16}}$ e $\mathbf{12_8 = \boldsymbol{A}_{16}}$ quindi $[\,2\,5\,3\,]_8 \rightarrow [\,\boldsymbol{A}\,\boldsymbol{B}\,]_{16}$

- Anche in questo caso la difficoltà di operare con basi diverse da 10 fa si che questo metodo sia utilizzato solo quando la base di partenza è 10

Divisioni ripetute – esempio 3

- Esempio 3: Convertire $[\mathbf{3323}]_{\mathbf{10}}$ in base 16

		16	← Base
$Q^{(0)}$	3323	11	← β_0
$Q^{(1)}$	207	15	← β_1
$Q^{(2)}$	12	12	← β_2
$Q^{(3)}$	0		

- In questo caso è necessario convertire i coefficienti $[\mathbf{12\ 15\ 11}]$ in base 16:
- $\mathbf{12_{10} = \mathit{C}_{16}, 15_{10} = \mathit{F}_{16}, 11_{10} = \mathit{B}_{16}}$
- Quindi $[\mathbf{3323}]_{\mathbf{10}} = \boldsymbol{CFA}_{\mathbf{16}}$

Conversioni fra basi diverse da 10

- Spesso si utilizza la base 10 come appoggio per convertire due basi differenti:
- Esempio 4:
  - convertire $[124]_5 = [\,?\,]_2$
  - $124_5 =$
  - $= 1 \cdot 5^2 + 2 \cdot 5^1 + 4 \cdot 5^0 =$
  - $= 25 + 10 + 4 =$
  - $= 39_{10}$
  - $124_5 = 39_{10} = 100111_2$

		2	← Base
$Q^{(0)}$	39	1	← β_0
$Q^{(1)}$	19	1	← β_1
$Q^{(2)}$	9	1	← β_2
$Q^{(3)}$	4	0	← β_3
$Q^{(4)}$	2	0	← β_4
$Q^{(5)}$	1	1	← β_5
$Q^{(6)}$	0		

Conversioni fra basi una potenza dell'altra

- In alcuni casi il passaggio di base fra due basi diverse da 10 è particolarmente semplice:
- Da base 2 a base 8, essendo $\mathbf{8 = 2^3}$

0 1 0	1 0 0	1 1 0	Base 2
↓	↓	↓	
2	4	6	Base 8

- Da base 2 a base 16, essendo 16= 2^4

1 0 1 0	0 1 1 0	Base 2
↓	↓	
A	6	Base 16

Conversioni fra basi una potenza dell'altra -2

- Da base 8 a base 2

3	2	6	Base 8
↓	↓	↓	
0 1 1	0 1 0	1 1 0	Base 2

- Da base 16 a base 2

1	A	F	Base 16
↓	↓	↓	
0 0 0 1	1 0 1 0	1 1 1 1	Base 2

Rappresentazione dei Numeri Interi

- $\boldsymbol{I} = \{\dots -3, -2, -1, 0, 1, 2, 3, \dots\}$
- Una possibile soluzione
  - codifica naturale per il valore assoluto
  - un simbolo aggiuntivo per il segno
    - In base 10 si usano i simboli aggiuntivi $\{+, -\}$per il segno
- Non volendo usare simboli aggiuntivi dobbiamo ricorrere a codifiche particolari:
  - Modulo e segno
  - Complemento alla base
  - Complemento alla base diminuita
  - Forma polarizzata

Codifica in modulo e segno

- Consideriamo $X = [\alpha_{n-1}\ \alpha_{n-2}\ \ldots\ \alpha_0\ ]$
- Il simbolo α_{n-1} indica il segno
  - $\alpha_{n-1} = 0$ indica valori positivi
  - $\alpha_{n-1} = b - 1$ indica valori negativi
- I restanti $\alpha_{n-2}\ \ldots\ \alpha_0$ indicano il valore assoluto (modulo)
- La conversione di base si effettua convertendo prima il modulo e poi il segno

$$[\ -41\ ]_{10} = [\ ?\ ]_{2,MS}$$
$$[\ -41\ ]_{10} = [\ 101001\ ]_{2}$$
$$[\ -41\ ]_{10} = [\ 1101001\ ]_{2,MS}$$

Codifica in modulo e segno - 2

- Ricaviamo il valore decimale di $323_{4,MS}$
  - Il bit di segno è $3 = 4 - 1$, il valore è negativo
  - $23_4 = 2 \cdot 4 + 3 = 11$, modulo
  - $323_{4,MS} = -11_{10}$
- Questa codifica è univoca?
- Ricaviamo il valore decimale di $00_{2,MS}$
  - Bit di segno: 0, valore positivo - Modulo: 0
  - $00_{2,MS} = 0_{10}$
- Ricaviamo il valore decimale di $10_{2,MS}$
  - Bit di segno: 1, valore negativo - Modulo: 0
  - $10_{2,MS} = -0_{10} = 0_{10}$
- Due rappresentazioni per lo 0, la codifica è ambigua
- Quanti numeri posso rappresentare con n simboli?
- $[-b^{n-1} + 1; +b^{n-1} - 1]$

Codifica in complemento alla base b

- Questa rappresentazione ha lo scopo di integrare l'informazione relativa al segno all'interno della parola di codice
  - Il segno non è indicato dalla cifra più a sinistra come in MS
  - **Il numero** n di simboli deve essere fissato a priori
- Si opera in modulo $\mathbf{b}^n$, infatti:
  - I valori positivi hanno codifica coincidente con la codifica naturale
  - Se il valore è negativo
  - $(-x)_{b,Cb} = b^n - x_{b,Cb}$
  - $x_{b,Cb} + (-x)_{b,Cb} = b^n = 0$
- Il valore 0 ha una sola codifica, la codifica è univoca!

Codifica in complemento alla base b – cont.

- Quanti simboli posso rappresentare con n cifre?
  - b^n parole di lunghezza n
  - una codifica per lo zero
- Ci sono $b^n - 1$ parole per codificare valori positivi e negativi
  - Se $b^n - 1$ è pari (b è dispari)
    - $\left[-\frac{b^n-1}{2}; +\frac{b^n-1}{2}\right]$
  - Se $b^n - 1$ è dispari (b è pari)
    - $\left[-\frac{b^n}{2}; +\frac{b^n}{2} - 1\right]$

Codifica in complemento alla base b - esempi

- Esempio 1
  - $[\mathbf{12}]_{\mathbf{10}} = [\,?\,]_{\mathbf{3,C3}}$
  - Essendo 12 un numero positivo, la sua codifica coincide con la codifica del valore naturale in base 3
  - $[\mathbf{12}]_{\mathbf{10}} = [\,\mathbf{110}\,]_{\mathbf{3,C3}}$
  - Esempio 2
  - $[\mathbf{-58}]_{\mathbf{10}} = [\,?\,]_{\mathbf{5,C5}}$
  - Si codifica il valore assoluto
  - $\mathbf{58_{10} = 213_5 = 213_{5,C5}}$
  - Quindi si cambia il segno
  - $\left(\mathbf{5^3}\right)_{\mathbf{5,C5}} - \mathbf{213_{5,C5} = 1000_{5,C5} - 213_{5,C5} = 232_{5,C5}}$

Complemento a 2

- Di particolare rilievo nell'ambito del corso
- Esempio: $[-24]_{10} = [\,?\,]_{2,C2}$, su 6 bit
  - 1° metodo
    - $24_{10} = 011000_2 = 011000_{2,C2}$
      $(2^6)_2 - 011000_{2,C2} = 1000000_{2,C2} - 011000_{2,C2} = 101000_{2,C2}$
  - 2° metodo (per evitare di svolgere la sottrazione in base 2)
    - Si complementano tutti i bit di 011000_2 e si somma il valore 1
      $$100111_2 + 000001_2 = 101000_{2,C2}$$
  - 3° metodo (si evita anche la somma in base 2)
    - Si scorre la parola da destra a sinistra lasciando inalterati tutti i bit fino al primo 1 e complementando i successivi
- In complemento a 2 si verifica che il bit più a sinistra fornisce un'indicazione del segno (0 positivo, 1 negativo)

Codifica in complemento alla base diminuita b-1

- Il segno è integrato nella rappresentazione, cambio di segno:
  - $(-x)_{b,Cb-1} = b^n - 1 - x_{b,Cb-1}$
  - Con n numero di simboli della parola di codice
  - Come in per il complemento a b si opera in modulo b^n

- Due rappresentazioni per lo zero: $\mathbf{0,\ b^n - 1}$
- $x_{b,Cb-1} + (-x)_{b,Cb-1} = x_{b,Cb-1} + b^n - 1 - x_{b,Cb-1} = b^n - 1$
  - Restano $b^n - 2$ simboli per rappresentare gli altri numeri
  - $\left[-\frac{b^n-2}{2}; +\frac{b^n-2}{2}\right]$
- La codifica dei numeri positivi coincide con quella in complemento a b che coincide con quella dei naturali
- La codifica cambia solo per i numeri negativi

Codifica in complemento a b-1 - esempi

- $[-45]_{10} = [\,?\,]_{4,C3} \qquad n = 4$

		4	← Base
$Q^{(0)}$	45	1	← β_0
$Q^{(1)}$	11	3	← β_1
$Q^{(2)}$	2	2	← β_2
$Q^{(3)}$	0		

- $45_{10} = 0231_{4,C3} \qquad n = 4$
- $-45_{10} = b^n - 1 - 0231_{4,C3} =$
- $-45_{10} = 3333_{4,C3} - 0231_{4,C3} = 3102_{4,C3}$

Codifica in complemento a 1

- Caso di particolare interesse, Base = 2
- Per cambiare segno ad un valore:
  - si sottrae il valore binario dalla parola 11…1
  - Quindi si fa il complemento logico (o negazione) di ogni bit
- Per questo motivo
  - Le codifiche che iniziano con 0 indicano valori positivi
  - Le codifiche che iniziano con 1 indicano valori negativi

- Esempio
  - $[-9_{10}] = [\,?\,]_{2,C1} \quad \boldsymbol{bit = 5}$
  - $\mathbf{9_{10} = 01001_{2,C1}}$
  - $\mathbf{-9_{10} = 10110_{2,C1}}$
  - È bastato fare il complemento logico di $\mathbf{01001}$

Codifica in forma polarizzata (eccesso-m)

- Il segno è integrato nella rappresentazione
- Codifica il valore
- $x \in [-m; +m+1]$
- codificando in codifica naturale il valore
- $x + m$
- Con $m = b^{n-1} - 1$ fattore di polarizzazione su n cifre

- Una sola rappresentazione per lo 0
- Per costruzione, al contrario delle codifiche precedenti
  - Le codifiche che iniziano con 0 indicano valori negativi
  - Le codifiche che iniziano con 1 indicano valori positivi

Codifica in forma polarizzata (eccesso-m)

- Esempio 1 (su 4 bit in base 2)
  - $6_{10} = [\,?\,]_{2,P}$
  - $\left(6 + \left(2^3 - 1\right)\right)_{10} = 13_{10} = 1101_2$
  - $6_{10} = 1101_{2,P}$

- Esempio 2 (su 4 bit in base 2)
  - $-6_{10} = [\,?\,]_{2,P}$
  - $\left(-6 + \left(2^3 - 1\right)\right)_{10} = 1_{10} = 0001_2$
  - $-6_{10} = 0001_{2,P}$