Codici di sorgente

$$C = (\mathcal{M}, \varphi, \mathcal{W}) \qquad \mathcal{M} \subseteq \mathcal{A}^+, \quad \mathcal{W} \subseteq \mathcal{B}^+, \quad \varphi : \mathcal{A}^+ \to \mathcal{B}^+$$

- Blocco - blocco (B-B)
- Blocco - lungh. variabile (B-LV)
- Lungh. variabile - blocco (LV-B)
- Lungh. variabile - lungh. variabile (LV-LV)

Esempi:

B-B: $\mathcal{M} = \mathcal{A} = \{a, b, c\}$, $\mathcal{W} = \{\varphi(a), \varphi(b), \varphi(c)\} = \{00, 01, 10\}$

$\mathcal{B} = \{0, 1\}$

$$\varphi(bbacab) = 01/01/00/10/00/01 = 010100100001$$

B-LV: $\mathcal{M} = \mathcal{A} = \{a, b, c\}$, $\mathcal{W} = \{\varphi(a), \varphi(b), \varphi(c)\} = \{0, 10, 11\}$

$$\varphi(bbacab) = 10/10/0/11/0/10 = 1010011010$$

LV-B: $\mathcal{M} = \{aa, ab, ac, b, c\}$, $\mathcal{W} = \{\varphi(aa), \varphi(ab), \dots\} = \{000, 001, 010, 011, 100\}$

$$\varphi(bbacab) = \varphi(b)\varphi(b)\varphi(ac)\varphi(ab) = 011/011/010/001 = 011011010001$$

LV-LV : $\mathcal{M}$ come sopra , $\mathcal{W} = \{\varphi(aa), \varphi(ab), \varphi(ac), \varphi(b), \varphi(c)\}$
$= \{000, 001, 01, 10, 11\}$

$\varphi(\underbrace{b}\underbrace{b}\underbrace{ac}\underbrace{ab}) = 10/10/01/001 = 101001001$

Univoca decodificabilità (u.d.)

Nel caso di codici con parole di codice a lunghezza costante (B-B o LV-B), per avere u.d.

è necessario e sufficiente che tutte le parole di codice siano distinte (φ è iniettiva)

Per codici con parole di codice a lunghezza variabile, ciò non è vero!

Esempio. $\mathcal{M} = \mathcal{A} = \{a, b, c\}$ $\mathcal{W} = \{\varphi(a), \varphi(b), \varphi(c)\} = \{0, 01, 001\}$

La sequenza secondaria 000101 si può decodificare sia $0/0/01/01 \to aabb$

sia $0/001/01 \to acb$

Ritardo di decodifica.

Esempio. $\mathcal{M} = \mathcal{A} = \{a, b, c\}$, $W = \{\varphi(a), \varphi(b), \varphi(c)\} = \{0, 01, 011\}$

Se ricevo 001010, l'unica decodifica possibile è 0/01/01/0 $\rightarrow$ abba

Dopo il 1° "0" ancora non so la parola di codice da decodificare è 0 oppure 01

(ritardo di decodifica pari a 1)

Esistono codici u.d. "patologici" con ritardi di decodifica arbitrariamente lunghi.

Esempio. $\mathcal{M} = \mathcal{A} = \{a, b, c\}$, $W = \{\varphi(a), \varphi(b), \varphi(c)\} = \{1, 10, 00\}$

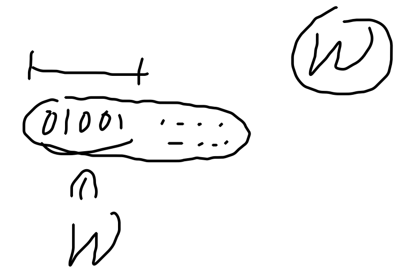
La sequenza $10^n1 = 1\underbrace{00\cdots0}_{n}1$ decodifica in $\begin{cases} ac^{n/2}a & \text{se } n \text{ è pari} \\ bc^{(n-1)/2}a & \text{se } n \text{ è dispari} \end{cases}$

(ritardo è proporzionale a n)

Un codice senza ritardo è detto istantaneo.

Def. Una famiglia di parole di codice $W = \{w_1, \dots, w_T\}$ è detta <u>a prefisso</u> se nessuna parola di codice è prefisso di un'altra.

(Codici a prefisso) (Esempio: prefissi telefonici) 06 3
063



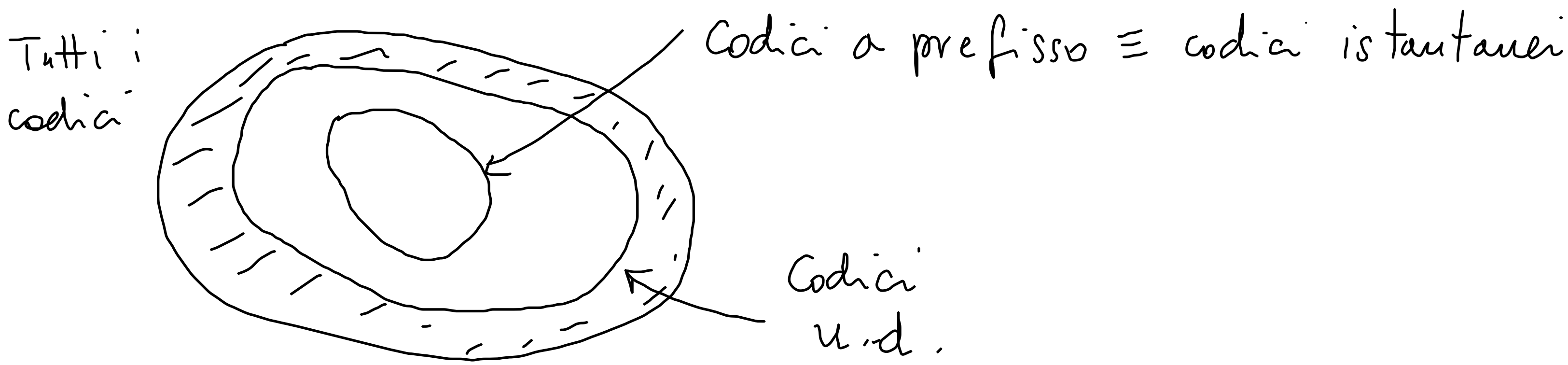
Prop. 3.1. Codice è a prefisso $\iff$ è istantaneo

Dim. $\Rightarrow$) evidente

$\Leftarrow$) Sia per assurdo $\alpha, \beta \in W$ con $\hat{\beta} = \hat{\alpha}\hat{\rho}$ (α è prefisso di β)

Dopo aver ricevuto la seq. di simboli α non so ancora se

α decodificare α o aspettare. Quindi il codice non è istantaneo. QED.



Il dizionario W di un codice può essere descritto attraverso un albero radicato (un grafo connesso aciclico, con un nodo radice evidenziato)

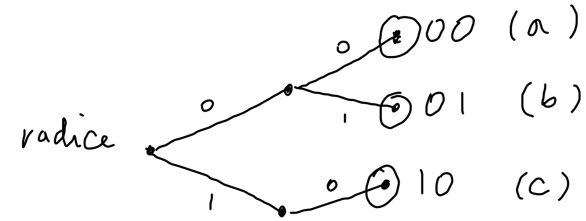
Alfabeto (secondario) con D simboli $\Rightarrow$ albero D-ario: al più D rami uscenti da ogni nodo. (Nodi senza rami uscenti sono detti terminali o foglie).

Il nodo associato alla parola di codice $\alpha = b_{i_1} b_{i_2} \cdots b_{i_n}$ è il nodo finale del cammino $b_{i_1}, b_{i_2}, \ldots, b_{i_n}$ che parte dalla radice

Esempio. $\mathcal{A} = \mathcal{M} = \{a, b, c\}$

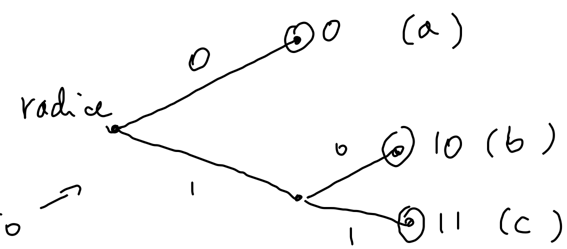
$\mathcal{B} = \{0, 1\}$ $\quad D = 2$

$W = \{\varphi(a), \varphi(b), \varphi(c)\} = \{00, 01, 10\}$



Non è completo

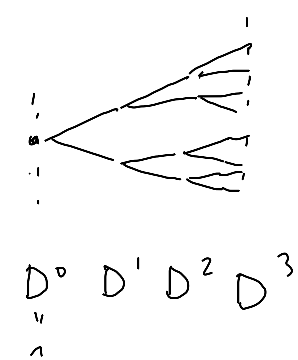
$W = \{\varphi(a), \varphi(b), \varphi(c)\} = \{0, 10, 11\}$



È completo

Albero completo: da ciascun nodo escono 0 oppure D rami

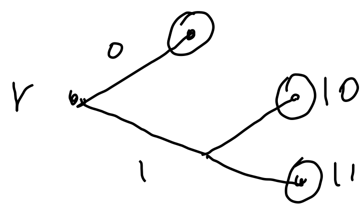
Albero zeppo: albero completo con D^n nodi di ordine n.



Proprietà del prefisso $\iff$ nell'albero di codice si incontra sempre ≤ 1 parola di codice nel percorso dalla radice ad ogni altro nodo.

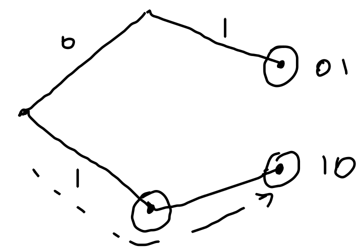
$D = |\mathcal{B}|$

Es. $W = \{0, 10, 11\}$



✓ È a prefisso

$W = \{01, 1, 10\}$



⊗ Non è a prefisso

Teo (3.9): Disuguaglianza di McMillan-Kraft: (*)

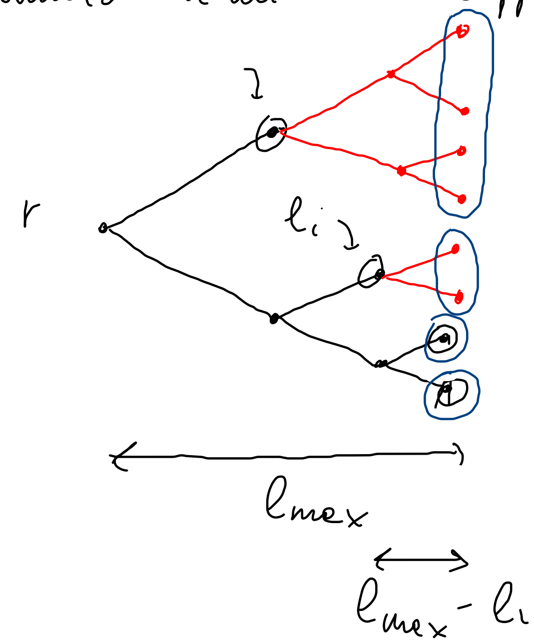
Se $l_1, l_2, \dots, l_K$ sono le lunghezze delle parole di un codice u.d., allora:

$$(*) \quad \sum_{i=1}^{K} D^{-l_i} \leq 1. \qquad (K = |\mathcal{M}|,\ D = |\mathcal{B}|)$$

Dim. Vediamo la dimostrazione solo nel caso particolare di codici a prefisso.

Consideriamo l'albero D-ario del codice ed estendiamolo ad un albero zeppo di altezza l_{max}.

$l_{max} = \max_{i=1}^{K} l_i$



$l_i = 0 \Rightarrow D^{l_{max}}$

$l_i = l_{max} \Rightarrow 1$

di discendenti di livello l_{max} di una parola di codice di lunghezza l_i è

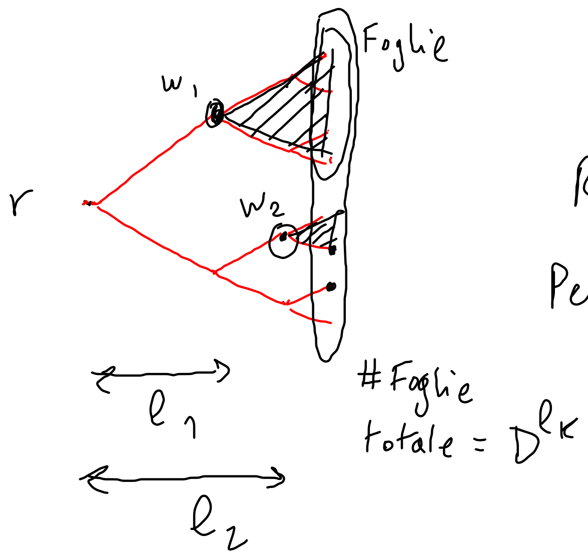
$D^{l_{max} - l_i}$

Se il codice è a prefisso, tutti questi insiemi di discendenti sono disgiunti. Ma abbiamo $D^{l_{max}}$ nodi terminali in totale (nell'albero esteso), quindi

$$\sum_{i=1}^{K} D^{l_{max} - l_i} \leq D^{l_{max}} \xRightarrow{\text{dividendo per } D^{l_{max}}} \sum_{i=1}^{K} D^{-l_i} \leq 1. \quad \text{QED.}$$

Teo (3.10): Se gli interi $D, \ell_1, \ell_2, \dots, \ell_K$ soddisfano la disuguaglianza $(*)$ allora esiste un codice D-ario u.d. (a prefisso) le cui parole $\{w_i\}$ hanno lunghezza ℓ_i.

Dim. Consideriamo l'albero zeppo di altezza $\ell_{max} = \max_{i=1}^{K} \ell_i$



Supponiamo (senza perdita di generalità) $\ell_1 \le \ell_2 \le \dots \le \ell_K = \ell_{max}$.

Rimuovo $D^{\ell_K - \ell_1}$ foglie

Per ipotesi $\sum_{i=1}^{K} D^{\ell_K - \ell_i} \le D^{\ell_K}$, quindi

$$D^{\ell_K - \ell_1} < D^{\ell_K}$$

$$D^{\ell_K - \ell_1} + D^{\ell_K - \ell_2} < D^{\ell_K}$$

$$\dots$$

Posso sempre trovare un percorso di lunghezza ℓ_i, per ogni $i = 1, 2, \dots, K$.

Il percorso di lunghezza ℓ_i identifica la parola di codice w_i. QED