



Teoria dell'Informazione e Applicazioni – a.a. 2014-2015

Esercizi su **Codifica** di Huffman e compressione

04-12-2014
Ing. P. Fazio

CODIFICA DI HUFFMAN

17

MOLTO UTILIZZATO PER COMPATTORE DATI (SPESSO USATO PER TESTO E BITMAP)

SI BASA SULLA COSTRUZIONE DELL'ALBERO: I CARATTERI PIÙ FREQUENTI SARANNO PIÙ VICINI ALLA RADICE, I MENO FREQUENTI PIÙ VICINI ALLE FOGLIE.

IN GENERALE SI RISPARMIA DAL 20% AL 90%

SI TRATTA DI CODICE PREFISSO, IN QUANTO NESSUNA PAROLA DI CODICE È PREFISSO DI UN'ALTRA (UTILE PER LA DECODIFICA, IN QUANTO OGNI PAROLA È UNIVOCAMENTE IDENTIFICATA)

ESEMPPIO CODICE PREFISSO: COSTRUZIONE ALBERO

0
101
100
111
1101
1100

PAROLE

CODICE:

001011101

0-0-101-1101

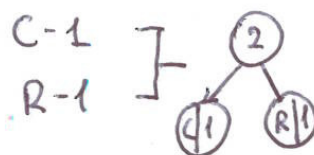
x y z w

A	3
L	2
C	1
R	1
M	1
E	1

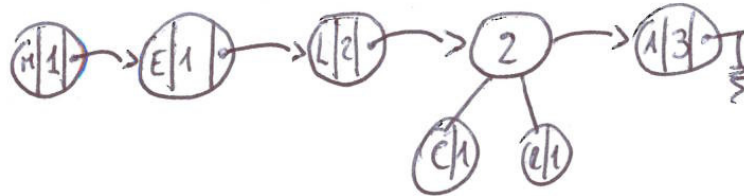
STEP 1): COSTRUZIONE DI UNA LISTA CONCATENATA IN ORDINE ASCENDENTE:



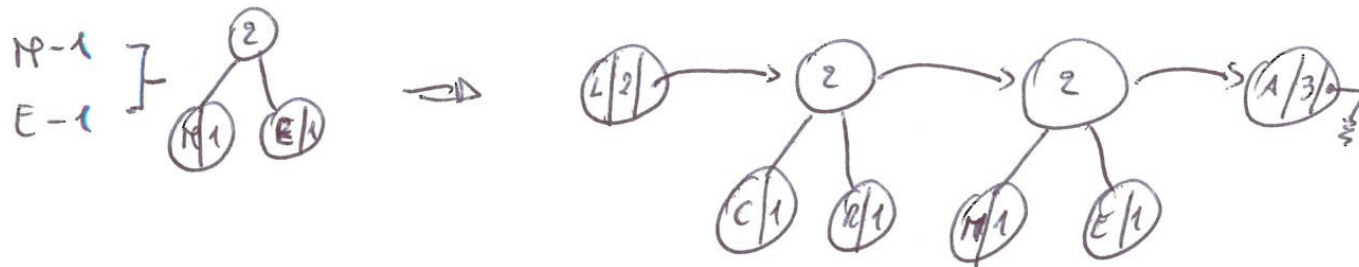
STEP 2) ANALISI DEI NODI A DUE A DUE: SI PARTE DAI PRIMI DUE (C, R)



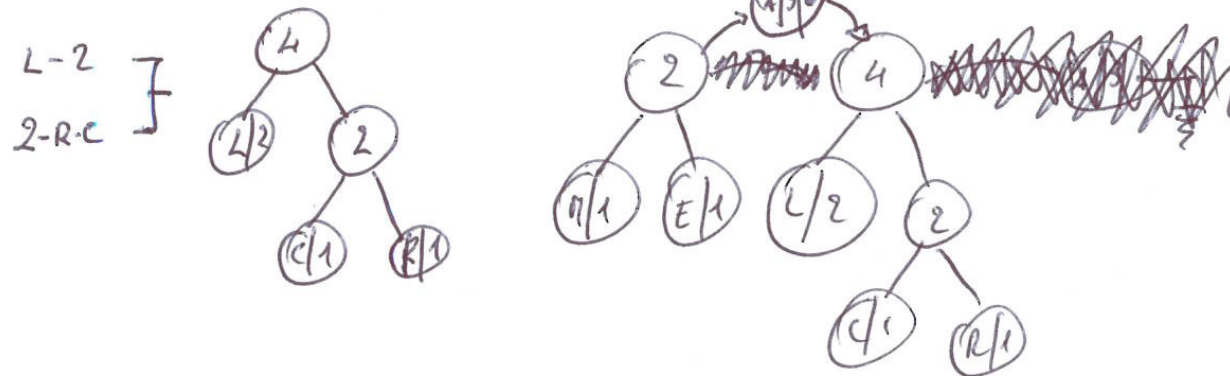
SI OTTIENE UN NUOVO NODO, IL CUI VALORE È DATO DALLA SOMMA DELLE FREQUENZE. SI OPERA UN INSERIMENTO ORDINATO, ELIMINANDO I NODI CONSIDERATI:



ADSSO SI PARTE IL CO STEP 2) PER I
NODI H E D E:



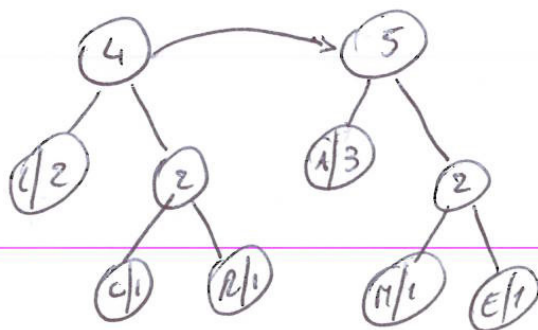
ADSSO SI HANNO 3 NODI CON "PESO" 2, SI PARTE SEMPLI DA SINISTRA:



SEMPRE LO STEP 2) PER I NODI 2-3

(18)

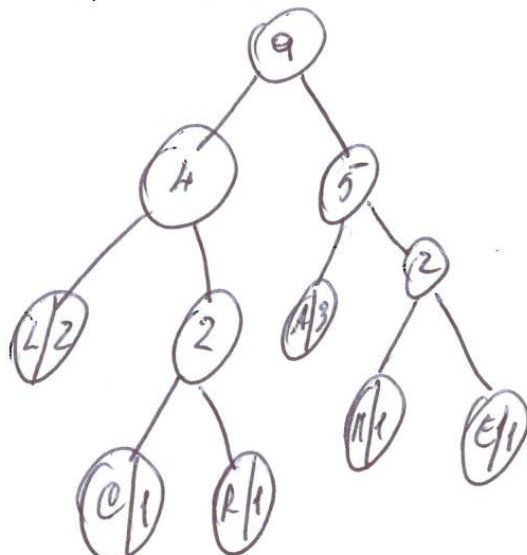
2
3 } 5



SEMPRE LO STEP 2) PER I NODI 4-5

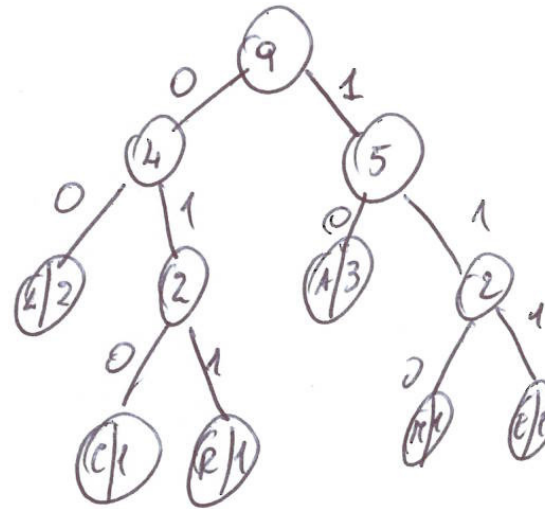
(LO STEP 2 TERMINA QUANDO SI OTTIENE UN UNICO NODO RADICE)

4
5 } 9



LETTERS
← ~~MA~~ PIÙ FREQUENTI SONO PIÙ VICINE ALLA ROOT (SI ASSIGNAHAN CODICI PIÙ BREVI)

STEP 3) ETICHETTARE I RAMI DEL SOTTOALBERO DA SINISTRA CON 0 E DI DESTRA CON 1:



AD OGNI LETTERA ASSOCIARE LA SEQUENZA DI BIT INCONTRATA PER ANDARE DALLA ROOT ALLA LETTERA

L → 00

A → 10

C → 010

R → 011

H → 110

E → 111

SE NOTA CHE AD OGNI LETTERA SI ASSOCIA UN NUMERO DI BIT MINORE DI 8.

DECODIFICARE:

010 10 011 10 110 111 00 00 10
 └─┘ └─┘ └─┘ └─┘ └─┘ └─┘ └─┘ └─┘
 C A R A H E L L A

ES. 2)

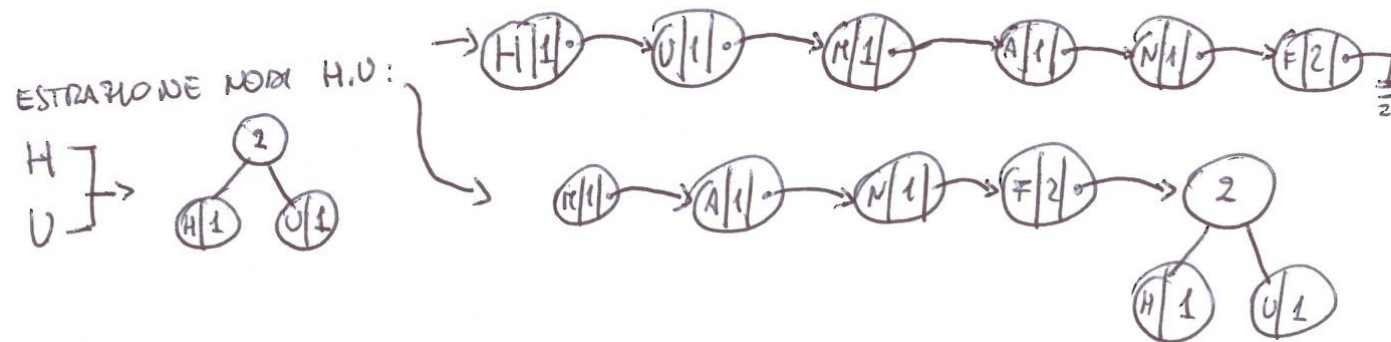
19

CALCOLARE LA % DI COMPRESSIONE PER LA PAROLA:
"HUFFMAN"

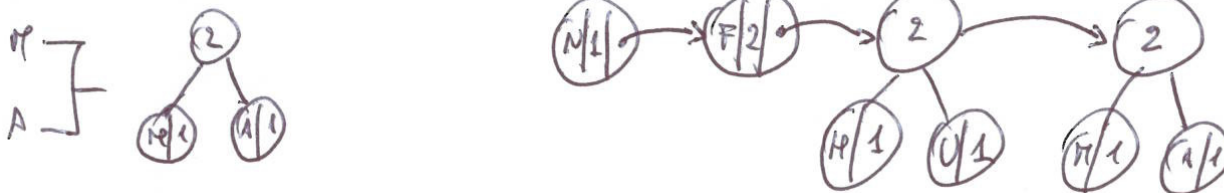
SE VIENE UTILIZZATA LA ~~PAROLA~~ CODIFICA DI HUFFMAN.

- SPAZIO OCCUPATO IN ORIGINE: 7 bytes (più 2 hanno 4 lettere)

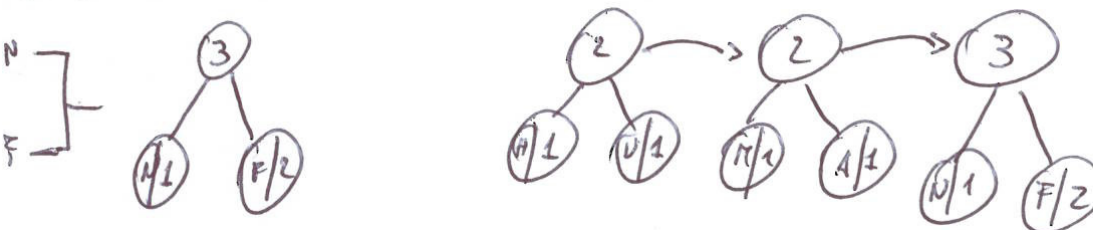
LA LISTA CONCATENATA DI PARTENZA È:



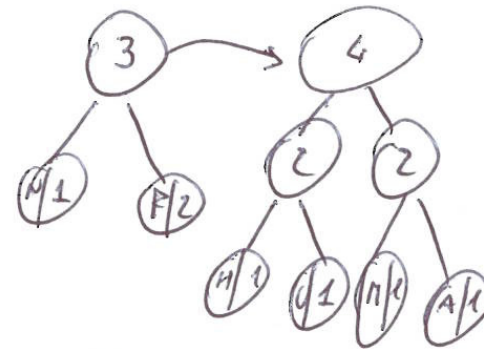
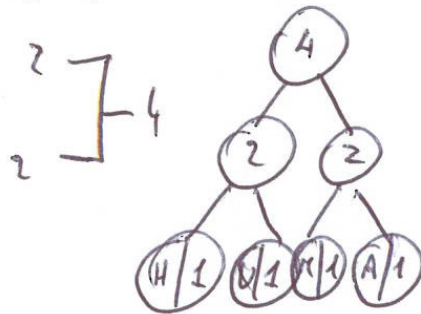
ESTRAZIONE NODI M, A:



ESTRAZIONE NODI N, F:

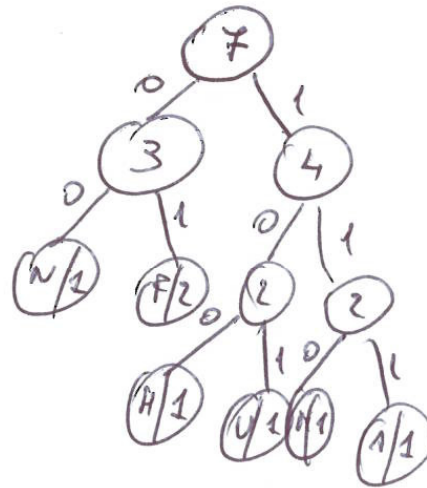


ESTRATTONS NOM 2,2:



ESTRATTONS NOM 3,4

3 } 7
4 }



N → 00

F → 01

H → 100

U → 101

M → 110

A → 111

LA PAROLA
CODIFICATA
SERÀ:

[10010101011011100]

18 bit, contro

56 bit in 2 di

$(1 - \frac{18}{56}) * 100 \approx 68\%$

ES. 3)

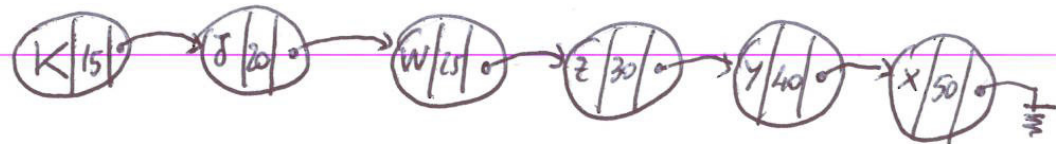
19

DATO UN FILE COMPOSTO DA 180 CARATTERI, CON FREQUENZE

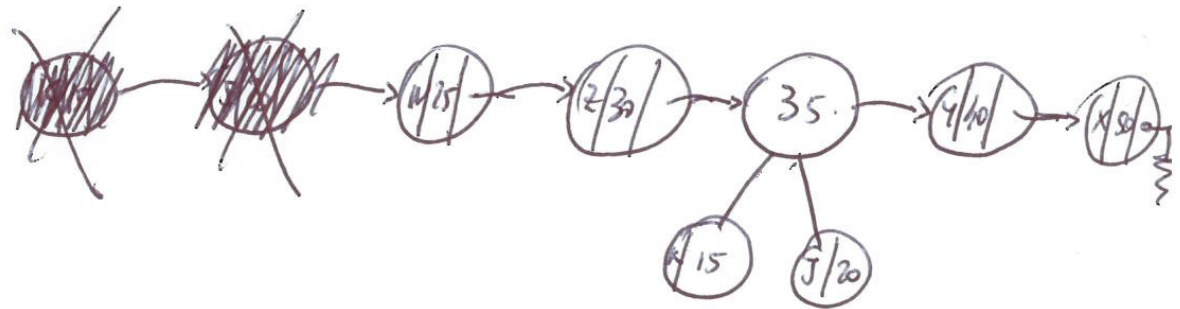
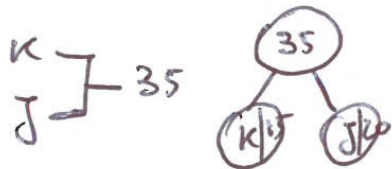
X Y Z W J K
50 40 30 25 20 15

LA CODIFICA DI
UTILIZZANDO HUFFMAN PER COMPRIMERE IL FILE

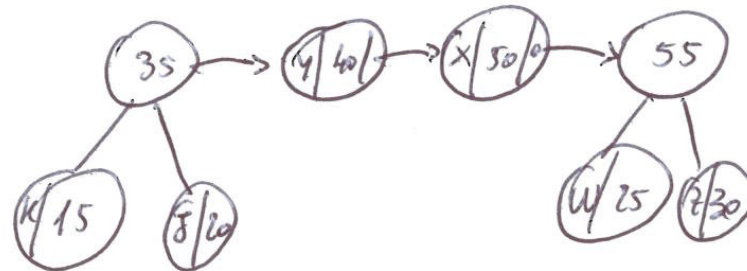
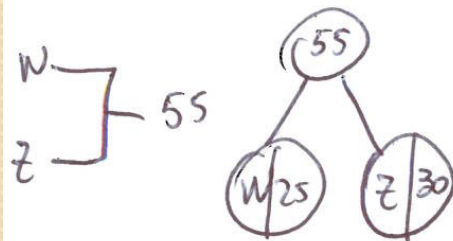
3.1) CREAZIONE LISTA:



3.2) ESTRAZIONE K, J

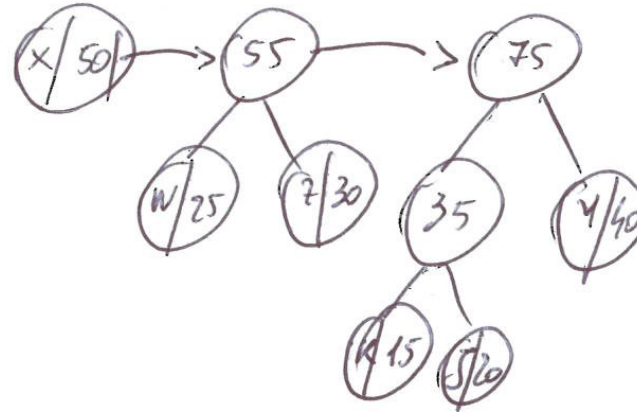
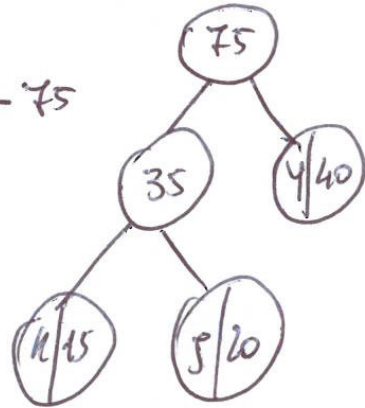


3.3) ESTRAZIONE W, Z:



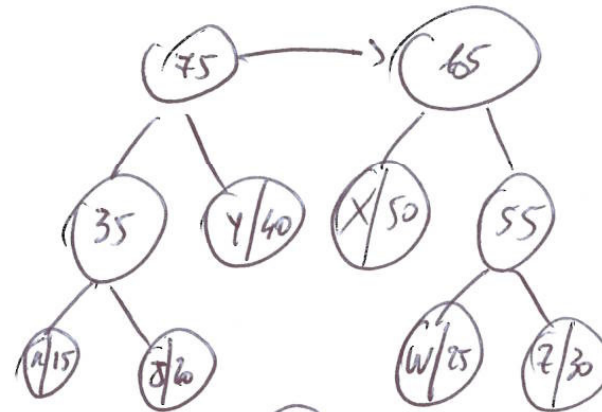
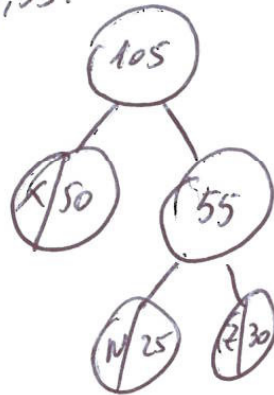
3.4) ESTRAZIONE 35, Y:

35 }
Y } 75



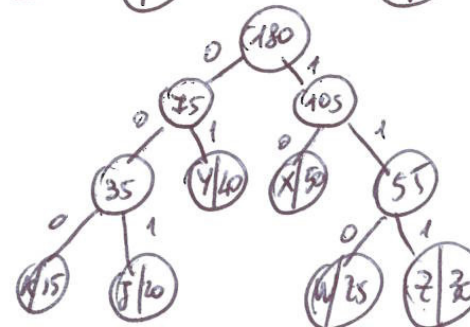
3.5) ESTRAZIONE 50, 55:

50 }
55 } 105



3.6) ESTRAZIONE 75, 105:

75 }
105 } 18



K → 000
J → 001
Y → 01
X → 10
W → 110
Z → 111

3.7) VALUTARE LA % DI COMPRESSIONE:

INIZIALMENTE ABBIAMO: $180 \cdot 8 = 1440 \text{ bit}$

CON HUFFMAN: $50 \cdot 2 + 40 \cdot 2 + 30 \cdot 3 + 25 \cdot 3 + 20 \cdot 3 + 15 \cdot 3 =$
 $x \quad y \quad z \quad w \quad j \quad k$

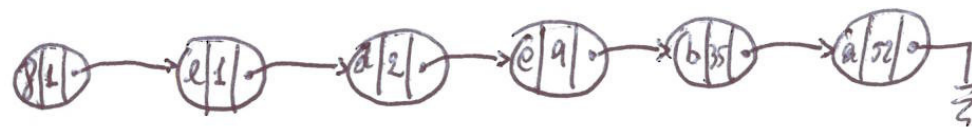
$$= 100 + 80 + 90 + 75 + 60 + 45 = \underline{\underline{450}}$$

$$C\% = \left(1 - \frac{450}{1440}\right) \times 100 \approx 68,7\%$$

~~3.8)~~

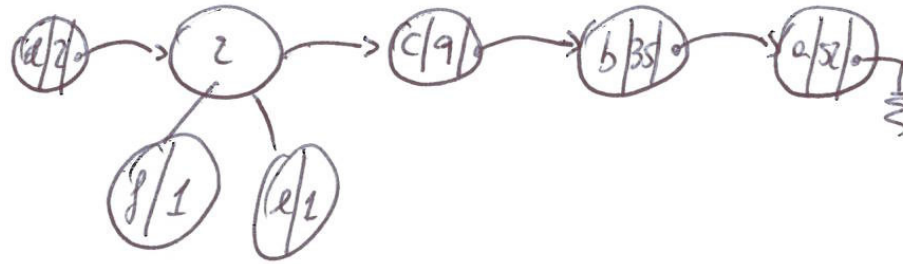
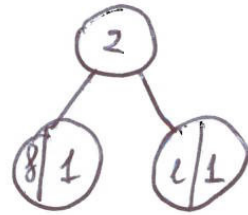
4) COSTRUIRE UN CODICE BINARIO OTTIMALE PER LA V.A. X , DEFINITA SULL'ALFABETO $\Omega = \{a, b, c, d, e, f\}$ CON DISTR. DI PROBABILITÀ $\{0,52, 0,35, 0,09, 0,02, 0,01, 0,01\}$.

APPLICANDO LA CODIFICA DI HUFFMAN AVREMO:



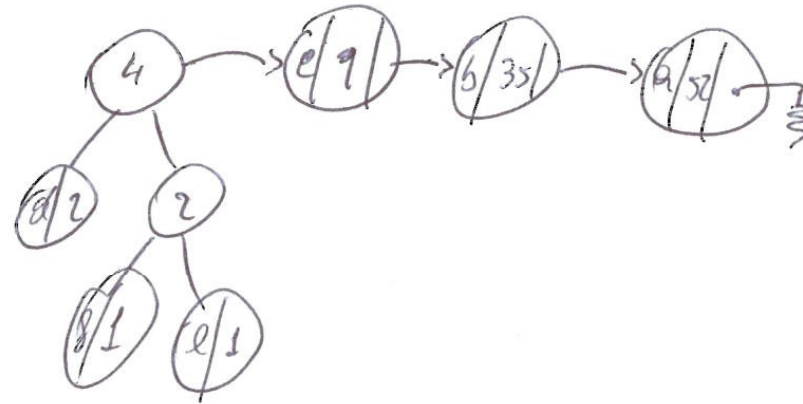
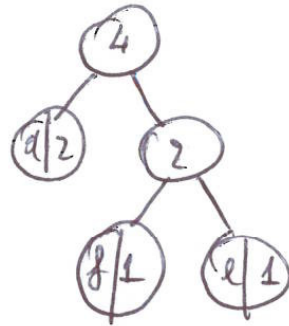
4.1) ESTIMATIONS f, e :

$\begin{bmatrix} f \\ e \end{bmatrix} = 2$



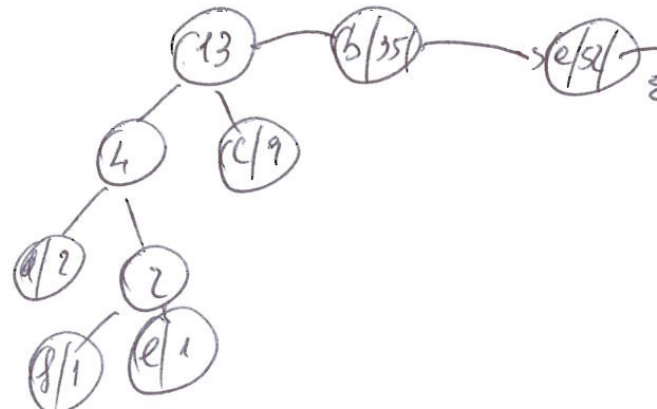
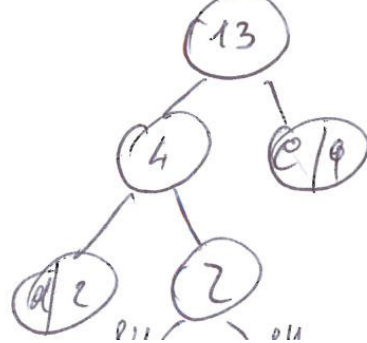
4.2) ESTIMATIONS d, 2:

$\begin{bmatrix} d \\ 2 \end{bmatrix} = 4$

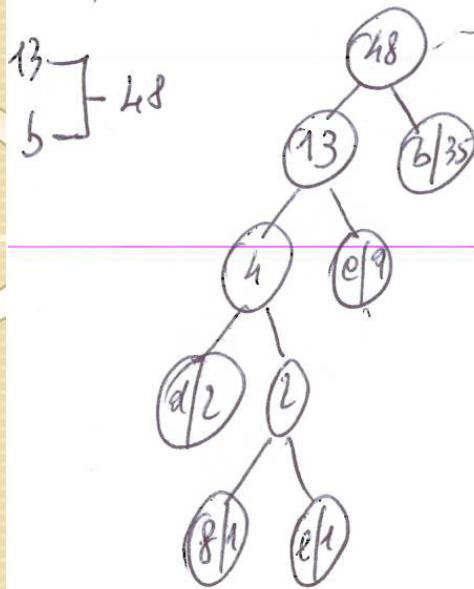


4.3) ESTIMATIONS 4, e

$\begin{bmatrix} 4 \\ e \end{bmatrix} = 13$

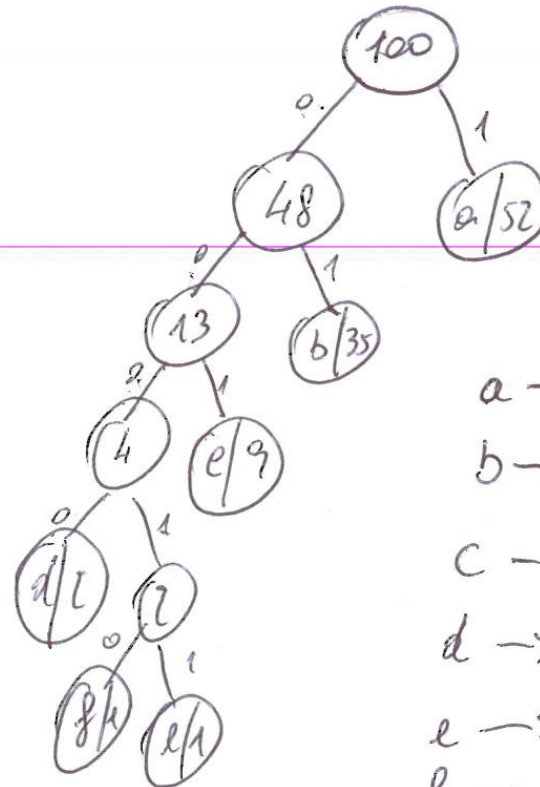


4.4) Estrazione 13, b



4.5) Estrazione 48, e (FINE)

(21)



$a \rightarrow 1$
 $b \rightarrow 01$
 $c \rightarrow 001$
 $d \rightarrow 0000$
 $e \rightarrow 00011$
 $f \rightarrow 00010$

4.5) CALCOLARE LA LUNGHEZZA MEDIA DEL CODICE

$$L(e) = \sum_{i=1}^6 p_i L_i = 0.52 \cdot 1 + 0.35 \cdot 2 + 0.09 \cdot 3 + 0.02 \cdot 4 + 0.01 \cdot 5 + 0.01 \cdot 5 = 0.52 + 0.70 + 0.27 + 0.08 + 0.05 + 0.05 = 1.67$$

\nwarrow prob. campione
 \nwarrow length code

4.6) CALCOLARE L'ENTROPIA ^{DI SHANNON} ~~DELLA~~ DELLA V.A.:

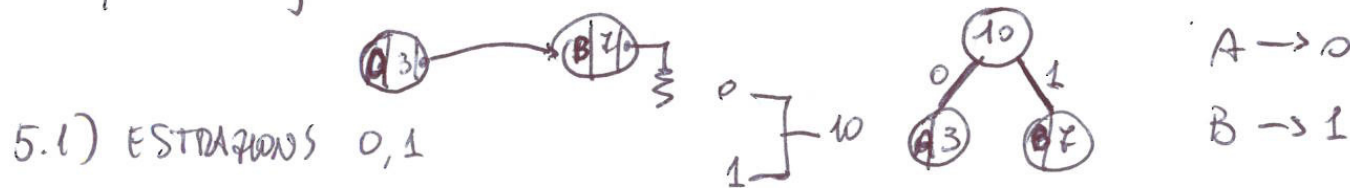
$$H(X) = \sum_{i=1}^6 p_i \log\left(\frac{1}{p_i}\right) = 0.52 \cdot 0.941 + 0.35 \cdot 1.511 + 0.09 \cdot 3.47 + 0.02 \cdot 5.644 + 0.01 \cdot 6.6$$

$$\approx 0.489 + 0.5285 + 0.3123 + 0.11288 + 0.1328 \approx 1.576$$

SI HA SEMPRE $H(X) < L(e)$ (È IL LOWER-BOUND PER $L(e)$) E RAPPRESENTA IL CONTENUTO INFORMATIVO MEDIO DELLA SORGENTE

IN BASE AL LOG UTILIZZATO (BASE 2, BASE E, BASE 10) IL RISULTATO È ESPRESSO IN BIT, NAT O HARTLEY. SE X AVESSSE UNA DISTR. D-ADICA ~~DELLA~~ ($\log_D\left(\frac{1}{p_i}\right) \in \mathbb{N}$), ALLORA $H(X) \equiv L(e)$

5) DATO UN ALFABETO (A, B) , SI TRASMETTONO I SIMBOLI CON PROBABILITÀ $P_X = \{0.3, 0.7\}$, CALCOLARE LUNGHEZZA MEDIA ^{ENTROPIA PER UNA} ~~DEL~~ CODICE DI HUFFMAN:



$$5.2) L(e) = \sum_{i=1}^2 p_i \cdot l_i = 0.3 \cdot 1 + 0.7 \cdot 1 = 1$$

(22)

$$H(x) = \sum_{i=1}^2 p_i \log_2 \left(\frac{1}{p_i} \right) = 0.3 \cdot 1.737 + 0.7 \cdot 0.5059 = 0.5211 + 0.35413 = 0.8752$$

5.3) CODIFICA DI HUFFMAN PER IL CODICE COMPOSTO $(A, B)^3$:

LE POSSIBILI PAROLE SONO: $\{AAA, AAB, ABA, ABB, BAA, BAB, BBA, BBB\}$, LE CUI

PROBABILITÀ SONO: $\{0.3^3, 0.3^2 \cdot 0.7, 0.3^2 \cdot 0.7, 0.3 \cdot 0.7^2, 0.7 \cdot 0.3^2, 0.7^2 \cdot 0.3, 0.7^2 \cdot 0.3, 0.7^3\}$ =

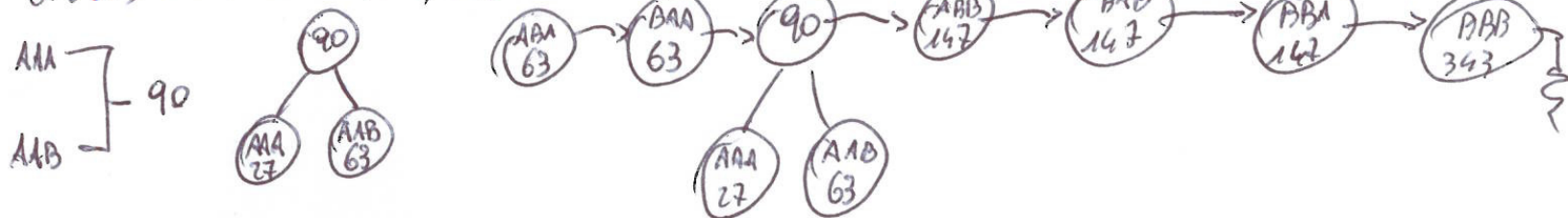
$$= \left\{ \overset{AAA}{0.027}, \overset{AAB}{0.063}, \overset{ABA}{0.063}, \overset{ABB}{0.147}, \overset{BAA}{0.063}, \overset{BAB}{0.147}, \overset{BBA}{0.147}, \overset{BBB}{0.343} \right\}$$

5.3.1) ORDINIAMO LE PAROLE IN BASE AI VALORI CRESCENTI DI PROBABILITÀ:

$$\left\{ \overset{AAA}{0.027}, \overset{AAB}{0.063}, \overset{ABA}{0.063}, \overset{BAA}{0.063}, \overset{ABB}{0.147}, \overset{BAB}{0.147}, \overset{BBA}{0.147}, \overset{BBB}{0.343} \right\}$$

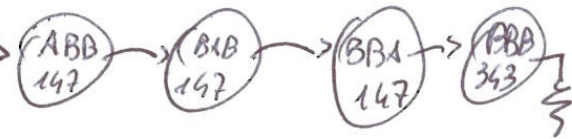
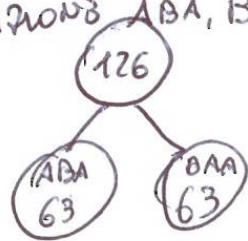


5.3.2) ESTRAGLIAMO AAA, AAB



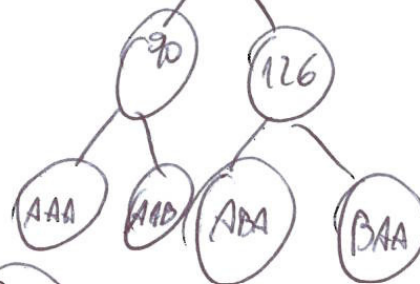
5.3.3) 8 STATIONS ABA, BAA

ABA
BAA } 126



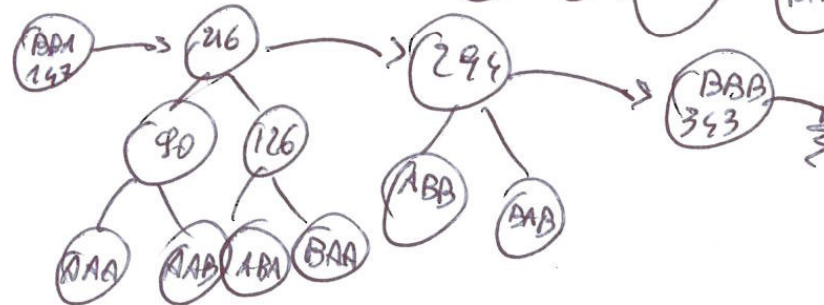
5.3.4) 8 STATIONS 90, 126

90
126 } 216

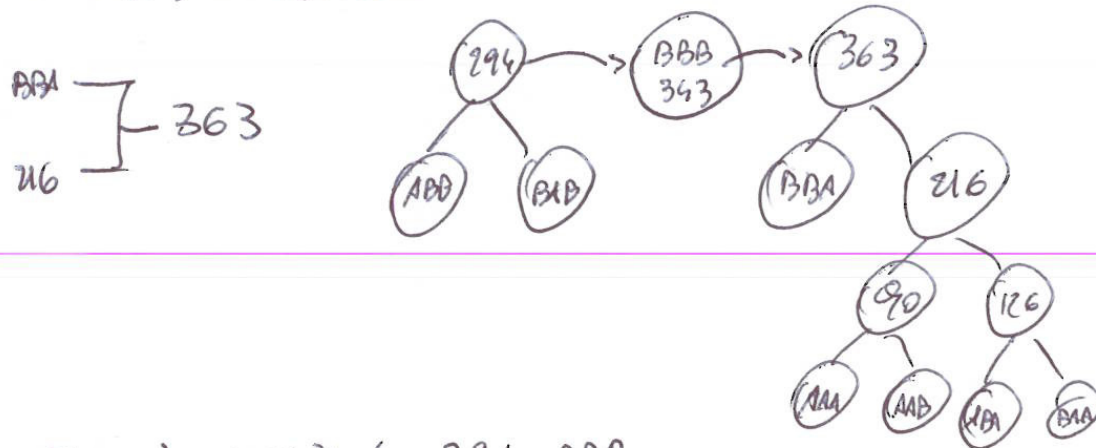


5.3.5) 8 STATIONS ABB, BAB

ABB
BAB } 294

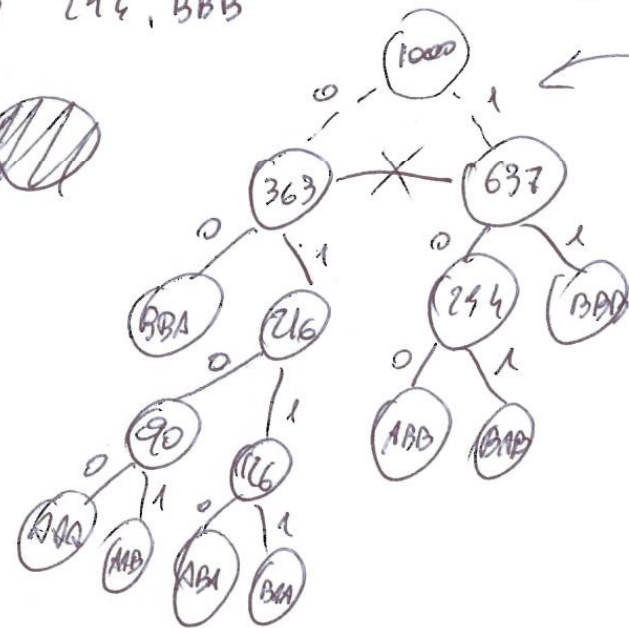


5.3.6) STRINGS BBA 216



5.3.7) STRINGS 294, BBB

294
BBB } 637



5.3.8) STRINGS 363, 637

- BBA → 00
- BBB → 11
- ABB → 100
- BAB → 101
- AAA → 0100
- AAB → 0101
- ABA → 0110
- BAA → 0111

PON CASA) VAWTAD L(c) & H(x).