

ΘΕΜΑ

Δίδεται κώδικας C που προέκυψε με εφαρμογή του αλγορίθμου κωδικοποίησης Huffman και του οποίου η αντιστοίχιση των κωδικών λέξεων με τα σύμβολα της πηγής (s1,...s9) περιέχεται στον ακόλουθο πίνακα:

Σύμβολα	Κωδικές Λέξεις
s1	011000
s2	01100100
s3	010
s4	0110011
s5	01101
s6	1
s7	01100101
s8	0111
s9	00

α) Να βρείτε τα σύμβολα εκείνα που έχουν το μεγαλύτερο και το μικρότερο πληροφοριακό περιεχόμενο. Εξηγήστε την απάντησή σας.

β) Αν δίνεται ότι τα σύμβολα της πηγής παράγονται βάσει των πιθανοτήτων {0.28, 0.27, 0.25, 0.1, 0.05, 0.04, 0.005, 0.003, 0.002 }, να αντιστοιχίσετε τις πιθανότητες αυτές στα ανωτέρω σύμβολα της πηγής, s1 έως s9, έτσι ώστε το μέσο μήκος των κωδικών λέξεων να παραμείνει βέλτιστο. Να εξηγήσετε την απάντησή σας.

Απάντηση

α) Το σύμβολο που μεταφέρει το μεγαλύτερο πληροφοριακό περιεχόμενο πρέπει να είναι ανάμεσα σε αυτά που αντιστοιχούν στις κωδικές λέξεις με το μεγαλύτερο μήκος αφού αυτές οι κωδικές λέξεις έχουν ανατεθεί στις μικρότερες πιθανότητες λόγω του αλγορίθμου του Huffman. Με μία επισκόπηση του πίνακα των κωδικών λέξεων παρατηρούμε ότι τα σύμβολα s2 και s7 έχουν τις μεγαλύτερες κωδικές λέξεις με μήκος 8. Κατά αντιστοιχία η s6 έχει το μικρότερο πληροφοριακό περιεχόμενο.

β) Για να αντιστοιχήσουμε τις πιθανότητες στα σύμβολα της πηγής θα πρέπει να δούμε τις αντίστοιχες κωδικές λέξεις. Με δεδομένο ότι οι κωδικές λέξεις προέκυψαν από τον αλγόριθμο Huffman θα πρέπει να ξεκινήσουμε από τα σύμβολα με τις δύο μικρότερες πιθανότητες και αυτές να τις αντιστοιχήσουμε στα σύμβολα με τις δύο μεγαλύτερες κωδικές λέξεις. Αυτό συμβαίνει διότι ο Huffman κατασκευάζει το δένδρο των κωδικών λέξεων από τα φύλλα προς την ρίζα, δηλαδή αρχίζει από τα σύμβολα με τις μικρότερες πιθανότητες και προχωράει προς τις μεγαλύτερες πιθανότητες. Άρα οι πιθανότητες 0.002 και 0.003 πρέπει να αντιστοιγηθούν στα σύμβολα s2 και s7 αφού αυτά έχουν τις μεγαλύτερες κωδικές λέξεις μήκους 8. Στη συνέχεια το σύμβολο s4 είναι το μοναδικό με μήκος κωδικής λέξης 7, συνεπώς θα πρέπει να αντιστοιγηθεί στην αμέσως μικρότερη πιθανότητα που είναι η 0.005. Συνεχίζοντας κατ' αυτό τον τρόπο μπορούμε να αντιστοιχήσουμε και τις υπόλοιπες πιθανότητες καταλήγοντας στον παρακάτω πίνακα.

Σύμβολα	Κωδικές λέξεις	Πιθανότητες
s6	1	0,28
s9	00	0,27
s3	010	0,25
s8	0111	0,1
s5	01101	0,05
s1	011000	0,04
s4	0110011	0,005
s7	01100101	0,003
s2	01100100	0,002

(Θέμα 4 ΕΞ2009Β)

ΘΕΜΑ - Στο αεροδρόμιο Ελευθέριος Βενιζέλος προσγειώνονται και απογειώνονται αεροπλάνα βάσει του ακόλουθου στατιστικού πίνακα όπου δίνονται ο τύπος και ο μέσος αριθμός αεροπλάνων ανά ώρα

Τύπος	Μέσος αριθμός αεροπλάνων ανά ώρα
Τύπος 1	9
Τύπος 2	5
Τύπος 3	3
Τύπος 4	2
Τύπος 5	1

Προκειμένου να καταγραφούν τα δεδομένα αυτά σε κεντρικό σύστημα, απαιτείται η κωδικοποίησή των στοιχείων σε δυαδικό κώδικα. Ο κώδικας αναφέρεται στον τύπο του αεροπλάνου κατά την άφιξη όπως αυτός ορίζεται παραπάνω.

Να χρησιμοποιήσετε για την κατασκευή του παραπάνω κώδικα τους αλγορίθμους

(α) Fano και

(β) Huffman

(γ) Να βρεθούν οι επιδόσεις των κωδίκων και να συγκριθούν.

(δ) Να κάνετε τους αντίστοιχους υπολογισμούς για κώδικα Shannon

ΑΠΑΝΤΗΣΗ

(α)

Αρχικά πρέπει να προσδιορίσουμε τον αριθμό αεροπλάνων ανά ώρα όπως μας ορίζει η άσκηση. Επομένως ο συνολικός αριθμός των αεροπλάνων που φθάνουν ανά ώρα στο συγκεκριμένο αεροδρόμιο είναι

$$9+5+3+2+1=20$$

(α). Για τον κώδικα Faπo θα έχουμε

Τύπος	Μέσος αριθμός αεροπλάνων ανά ώρα	Πιθανότητες
Τύπος 1	9	$9/20=0.45$
Τύπος 2	5	$5/20=0.25$
Τύπος 3	3	$3/20=0.15$
Τύπος 4	2	$2/20=0.10$
Τύπος 5	1	$1/20=0.05$

						Κωδική λέξη	
Τύπος 1	0.45	0				0	
Τύπος 2	0.25	1	0				10
Τύπος 3	0.15	1	1	0			110
Τύπος 4	0.10	1	1	1	0		1110
Τύπος 5	0.05	1	1	1	1		1111

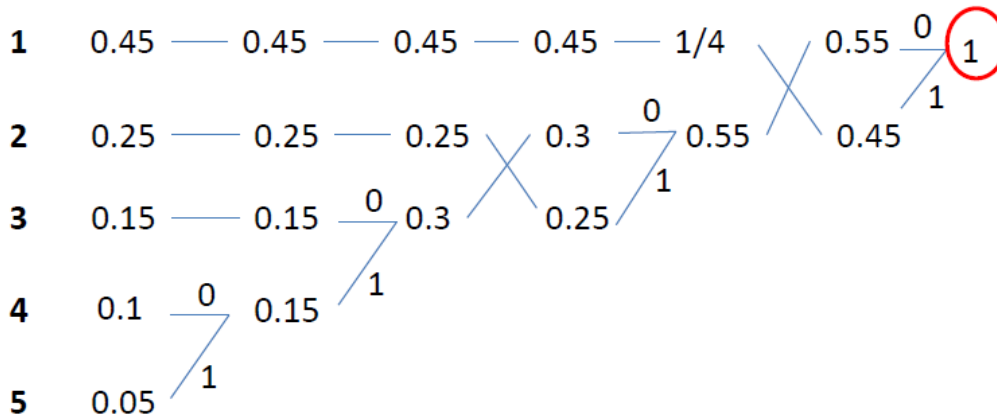
(β).

Για τον κώδικα Huffman θα έχουμε

Πιθανότητες Σύμβολα Συμβόλων		Βήμα 1	Βήμα 2	Βήμα 3	Βήμα 4	Κωδική Λέξη
Τύπος 1	0.45	0,45	0,45	0,45	0,55 (0)	1
Τύπος 2	0.25	0,25	0,25	0,30 (0)	0,45 (1)	01
Τύπος 3	0.15	0,15	0,15 (0)	0,25 (1)	0,45 (1)	000
Τύπος 4	0.10	0,10 (0)	0,15 (1)	0,25 (1)	0,45 (1)	0010
Τύπος 5	0.05	0,05 (1)	0,15 (1)	0,25 (1)	0,45 (1)	0011

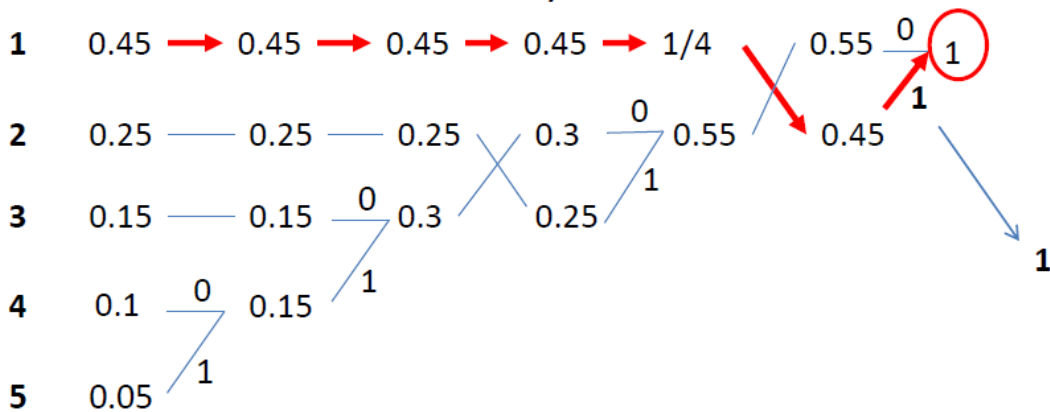
Παρακάτω παρατίθεται αναλυτικότερα ο αλγόριθμος κωδικοποίησης Huffman

Τύπος

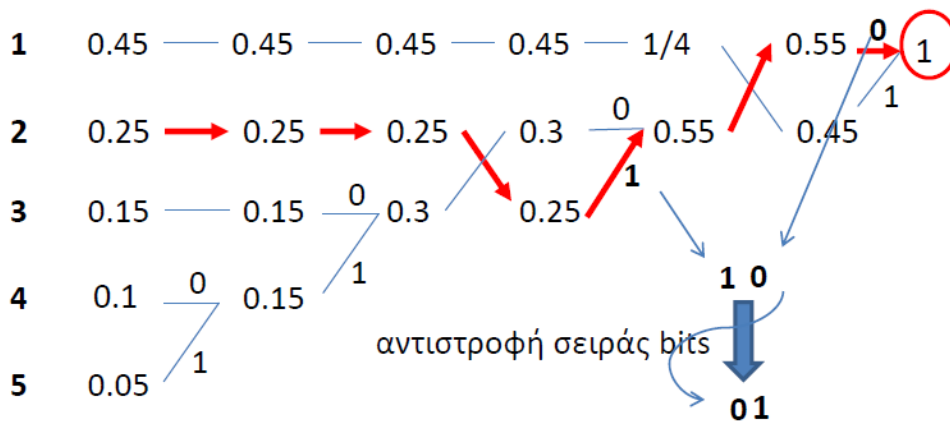


Υπολογισμός κωδικολέξεων για κάθε σύμβολο:

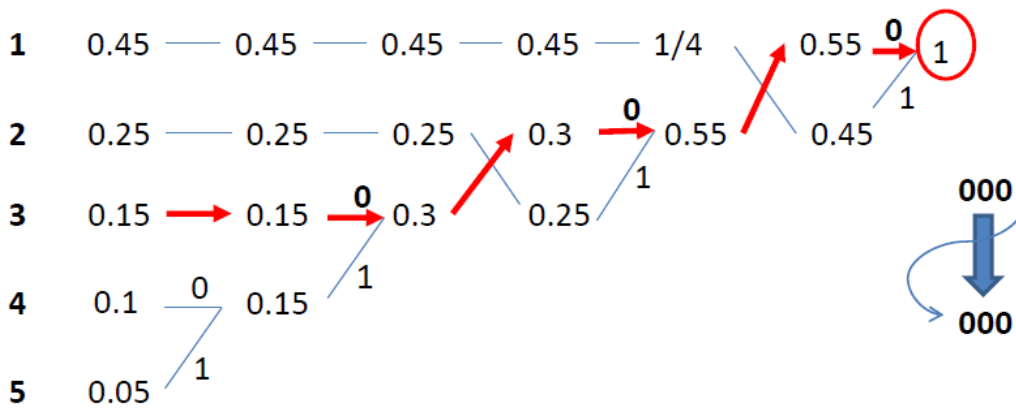
Τύπος 1



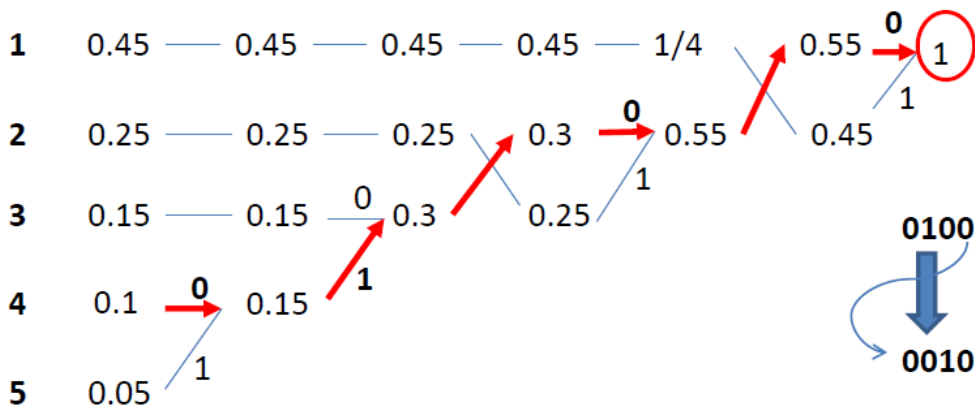
Τύπος 2



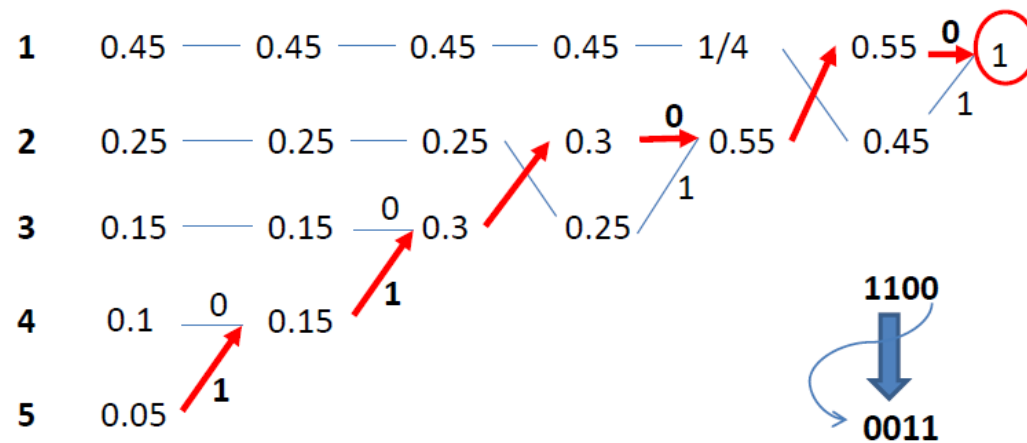
Τύπος 3



Τύπος 4



Τύπος 5



(γ)

Οι επιδόσεις ενός δυαδικού κώδικα δίνεται από τον τύπο 2.9, σελ. 57

$$a = \frac{H(C)}{\left(\sum_{i=1}^5 p_i l_i \right) \log_2 2}$$

Δηλαδή για τις επιδόσεις του κώδικα, υπολογίζουμε πρώτα την εντροπία της πηγής και ακολούθως υπολογίζουμε το μέσο μήκος των κωδικών λέξεων για κάθε περίπτωση και την αντίστοιχη απόδοση.

$$H(S) = -\sum_{i=1}^5 p_i \log p_i = -0.45 \log_2 (0.45) - 0.25 * \log_2 (0.25) - 0.15 * \log_2 (0.15) - 0.10 * \log_2 (0.10) - 0.05 * \log_2 (0.05) = 1.977 \text{ bits / symbol}$$

Κώδικας Fano

$$\sum_{i=1}^5 p_i l_i = 0.45 * 1 + 0.25 * 2 + 0.15 * 3 + 0.10 * 4 + 0.05 * 4 = 2$$

Εφαρμόζοντας τον προηγούμενο τύπο θα έχουμε

$$a_{Fano} = \frac{1.977}{2} = 0.9885$$

Κώδικας Huffman

$$\sum_{i=1}^5 p_i l_i = 0.45 * 1 + 0.25 * 2 + 0.15 * 3 + 0.10 * 4 + 0.05 * 4 = 2$$

Εφαρμόζοντας τον προηγούμενο τύπο θα έχουμε

$$a_{Huffman} = \frac{1.977}{2} = 0.9885$$

Κώδικας Shannon

Σύμβολα	ρi	πι	-log(ρi)	li	ρi x li
Τύπος 1	0,45	0	1,152003	2	0,9
Τύπος 2	0,25	0,45	2	2	0,5
Τύπος 3	0,15	0,7	2,736966	3	0,45
Τύπος 4	0,1	0,85	3,321928	4	0,4
Τύπος 5	0,05	0,95	4,321928	5	0,25
				μέσο μήκος	2,5

	π1		Ψηφίο
Τύπος 1	0	x 2 =	0
	0	x 2 =	0
			κωδ.λέξη 00

	π2		Ψηφίο
Τύπος 2	0,45	x 2 =	0,9
	0,9	x 2 =	1,8
			κωδ.λέξη 01

	π3		Ψηφίο
Τύπος 3	0,7	x 2 =	1,4
	0,4	x 2 =	0,8
	0,8	x 2 =	1,6
			κωδ.λέξη 101

	π4		Ψηφίο
Τύπος 4	0,85	x 2 =	1,7
	0,7	x 2 =	1,4
	0,4	x 2 =	0,8
	0,8	x 2 =	1,6
			κωδ.λέξη 1101

	π5		Ψηφίο	
Τύπος 5	0,95	x 2 =	1,9	1
	0,9	x 2 =	1,8	1
	0,8	x 2 =	1,6	1
	0,6	x 2 =	1,2	1
	0,2	x 2 =	0,4	0
			κωδ.λέξη	11110

Επίδοση $\alpha=1,97/2,5= 0,788$

(Θέμα 5 ΕΞ2007Β)