

4^η Σειρά ασκήσεων
(Διάφορα Θέματα)
 Ανάθεση: 6 Ιουνίου
 Παράδοση: 24 Ιουνίου

Άσκηση 1 (2 βαθμ.)

Θεωρείστε το αλφάβητο {α,β,γ,δ,ε} και την εξής φράση: «α α β γ γ α α α α α δ β ε β».

- α) Βάσει αυτής της φράσης ποια είναι η εντροπία του αλφαβήτου;
 β) Δώστε τη συμπίεσμένη μορφή της φράσης χρησιμοποιώντας κανονικοποιημένους κώδικες Huffman.

Άσκηση 2 (2 βαθμ.)

Θεωρείστε 8 έγγραφα A, B, Γ, Δ, E, Z, H, Θ, I, K και έστω ότι οι αποστάσεις μεταξύ τους είναι αυτές του παρακάτω πίνακα. Δώστε το δενδρικό διάγραμμα που προκύπτει εφαρμόζοντας ιεραρχική ομαδοποίηση εγγράφων τύπου: (α) SingleLink και (β) CompleteLink.

A										
B	1									
Γ	2	1								
Δ	3	2	1							
E	4	3	2	1						
Z	2	3	4	5	6					
H	3	2	3	4	5	1				
Θ	4	3	2	3	4	2	1			
I	5	4	3	2	3	3	2	1		
K	6	5	4	3	2	4	3	2	1	
	A	B	Γ	Δ	E	Z	H	Θ	I	K

Άσκηση 3 (3 βαθμοί)

Θεωρείστε τα ακόλουθα έγγραφα όπου τα γράμματα A-E συμβολίζουν λέξεις.

- d1 = «B Γ B », d2 = «B A A B»
 d3 = «A B », d4 = «Γ E Γ E»
 d5 = «Δ Γ Γ A», d6 = «Γ E»
 d7 = «B Δ B», d8 = «E B A»

Έστω ότι τα d1,d5, d6 ανήκουν σε ένα σύστημα S1, τα d2,d4 σε ένα σύστημα S2, και τα υπόλοιπα (d3,d7,d8) σε ένα σύστημα S3. Θέλουμε να φτιάξουμε έναν μεσίτη M πάνω από αυτά τα συστήματα.

(α) Για την επιλογή πηγής ο M θέλει να περιγράψει τα περιεχόμενα της κάθε πηγής με ένα διάνυσμα. Δώστε τα διανύσματα πηγών των S1, S2 και S3.

(β) Έστω ότι ο M έχει ήδη τα διανύσματα πηγών των S1, S2, S3 και λαμβάνει την επερώτηση q= «A Γ». Αν θέλει να προωθήσει την επερώτηση q σε μία μόνο πηγή, ποια θα επιλέξει;

(γ) Ο M λαμβάνει μια επερώτηση, την προωθεί σε όλες τις πηγές, και λαμβάνει τα εξής αποτελέσματα από την κάθε μια:

S1: <d5, d1, d6 >

S2: <d2, d4>

S3: <d7, d8, d3>

Δώστε την ενοποιημένη διάταξη κατά round robin interleaving

(δ) Προκειμένου ο μεσίτης να λαμβάνει από τις πηγές απαντήσεις με συγκρίσιμα σκορ, αποφασίζει να κάνει αποτίμηση επερωτήσεων σε δυο φάσεις ώστε οι πηγές να λαμβάνουν τα καθολικά στατιστικά που χρειάζονται για τον σωστό υπολογισμό των σκορ. Δώστε το idf του κάθε όρου στην καθολική συλλογή εγγράφων.

(ε) Ο μεσίτης βρίσκει άλλο ένα σύστημα S4 το οποίο έχει την ίδια συλλογή με αυτήν του S1, δηλαδή και αυτό παρέχει πρόσβαση στα έγγραφα d1, d5, d6. Έστω ότι ο M προωθεί μια επερώτηση q στα S1 και S4 και λαμβάνει τις εξής απαντήσεις:

S1: <d1, d5, d6>

S4: <d6, d5, d1>

Ποιο είναι το κορυφαίο έγγραφο αν ενοποιήσουμε τις διατάξεις: (i) κατά Borda, (ii) κατά Condorcet;

Ο M αποφασίζει να δίνει στο χρήστη όχι μόνο την ενοποιημένη διάταξη, αλλά και την Kemeny distance μεταξύ των διατάξεων που έλαβε από τα υποσυστήματα (προκειμένου ο χρήστης να παίρνει μια γεύση για το βαθμό συμφωνίας των πηγών). Ποια είναι αυτή η απόσταση στην προκειμένη;

(στ) Τα συστήματα S1, S2, S3 δεν θέλουν πλέον να έχουν ανάγκη τον M και αποφασίζουν να «ανεξαρτητοποιηθούν» φτιάχνοντας ένα σύστημα ομοτίμων (P2P), συγκεκριμένα ένα δομημένο σύστημα τύπου Chord. Προσελκύουν μάλιστα άλλα δυο συστήματα S5 και S6 (τα οποία δεν έχουν καμία συλλογή εγγράφων).

Αποφασίζουν να χρησιμοποιήσουν μια συνάρτηση κατακερματισμού h των 3 bits, και έστω ότι

$h(\text{IPaddress}(S1))=2$, $h(\text{IPaddress}(S2))=5$, $h(\text{IPaddress}(S3))=3$,
 $h(\text{IPaddress}(S5))=1$, $h(\text{IPaddress}(S6))=4$

Αποφασίζουν να διανείμουν το ανεστραμμένο ευρετήριο θεωρώντας κάθε όρο σαν κλειδί και έστω ότι

$h(A)=2$, $h(B)=3$, $h(\Gamma)=6$
 $h(\Delta)=6$, $h(E)=5$

Δώστε (i) τους πίνακες δρομολόγησης των κόμβων S1 και S3 και (ii) πως θα καταταξιολογηθεί το ανεστραμμένο ευρετήριο στους κόμβους του δικτύου (δείξτε τι ακριβώς θα έχει κάθε κόμβος)

Άσκηση 4 (3 βαθμοί)

Θέλουμε να ευρετηριάσουμε μια συλλογή εικόνων και μας δίνουν υλοποιημένη μια συνάρτηση απόστασης d η οποία δεν ικανοποιεί την τριγωνική ανισότητα. Θα μπορούσαμε σε αυτήν την περίπτωση να χρησιμοποιήσουμε μετρικά ευρετήρια; Αν όχι γιατί;