

ΑΝΑΦΟΡΑ ΚΑΙ ΣΧΟΛΙΑΣΜΟΣ ΜΕΤΡΗΣΕΩΝ

1. Περιγραφή κώδικα:

Η υλοποίηση του κώδικα της άσκησης πραγματοποιήθηκε στη γλώσσα προγραμματισμού python. Δημιουργήθηκαν οι ζητούμενες κλάσεις:

- **DataPage**: Χρησιμοποιήθηκε για να παραστήσει τις σελίδες του βασικού αρχείου στο `index_file` και το `sorted_index_file` (χρησιμοποιήθηκε για τον Γ τρόπο) ,για να αποθηκεύσει τα κλειδιά και να δημιουργηθούν τα απαραίτητα στιγμιότυπα που θα μεταφερθούν σε λίστες.
- **DataClass**: Χρησιμοποιήθηκε για να παραστήσει τα data και keys που αποθηκευόντουσαν και για να δημιουργηθούν χρήσιμα στιγμιότυπα σε λίστες

Ακόμα δημιουργήθηκαν και δύο μέθοδοι για την πιο εύκολη διαχείριση του κώδικα.

Η πρώτη ,με το όνομα `binary_search`, χρησιμοποιήθηκε στο να κάνει τη δυαδική αναζήτηση στον Γ τρόπο και μετέπειτα να επιστρέφει τα αποτελέσματα για να συνεχιστεί η `main` με τις τιμές επιστροφής να είναι οι ακόλουθες :

- Ο αριθμός εγγραφής και ο αριθμός σελίδας σε περίπτωση που βρέθηκε σε κάποιο σημείο
- -1 και ο αριθμός της σελίδας σε περίπτωση που δεν βρέθηκε ο αριθμός στο διάστημα
- -1 και το μέγεθος του αρχείου σε περίπτωση που δεν βρεθεί το διάστημα

Η δεύτερη , με όνομα `file_reader` χρησιμοποιήθηκε για να διαβάζουμε το περιεχόμενο του αρχείου και να συμπληρώνει τις λίστες που θα χρησιμοποιηθούν μετέπειτα για αναζήτηση.

Το υπόλοιπο του κώδικα ανήκει στη `main` και η επεξήγηση του γίνεται με σχόλια στο εσωτερικό του κώδικα.

2. Εξωτερικές πηγές

Για την δυαδική αναζήτηση χρησιμοποιήθηκε ένα σημείο κώδικα που τροποποιήθηκε για να καλύψει τις ανάγκες του προγράμματος που βρίσκεται στον παρακάτω σύνδεσμο:

<https://www.geeksforgeeks.org/python-program-for-binary-search/>

και για τη διαχείριση της θέσης του δείκτη του αρχείου τις πληροφορίες που παρείχε ο παρακάτω σύνδεσμος:

<https://pythonlobby.com/seek-and-tell-functions-in-python/>

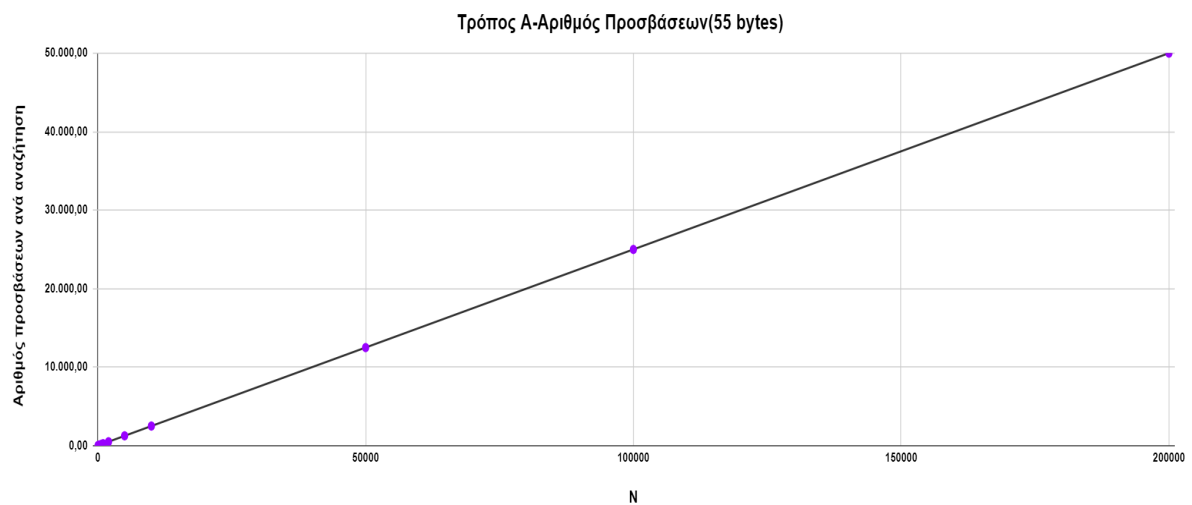
3. Σχολιασμός μετρήσεων Πίνακες

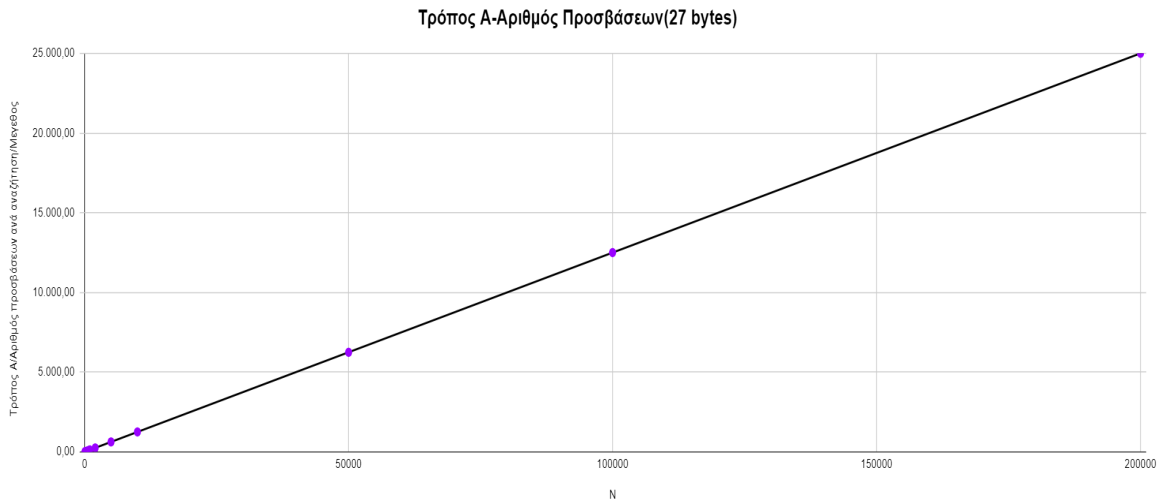
	Τρόπος Α		Τρόπος Β		Τρόπος Γ	
N	Αριθμός προσβάσεων ανά αναζήτηση		Αριθμός προσβάσεων ανά αναζήτηση		Αριθμός προσβάσεων ανά αναζήτηση	
	Μεγεθος δεδομένων 55 bytes	Μεγεθος δεδομένων 27 bytes	Μεγεθος δεδομένων 55 bytes	Μεγεθος δεδομένων 27 bytes	Μεγεθος δεδομένων 55 bytes	Μεγεθος δεδομένων 27 bytes
50	9,20	4,80	2,20	1,90	1,80	1,90
100	18,30	9,30	3,50	3,30	3,40	2,00
200	42,10	20,50	6,30	6,00	2,60	2,70
500	116,70	58,00	15,20	15,00	3,40	3,30
800	191,30	95,40	24,10	24,10	4,90	3,90
1000	240,40	119,10	31,00	30,70	4,30	5,10
2000	490,00	245,50	61,90	62,10	5,10	5,10
5000	1.245,40	621,60	156,60	156,40	6,40	6,40
10000	2.491,50	1.247,50	312,10	312,60	7,40	7,50
50000	12.486,60	6.244,00	1.561,50	1.561,80	9,60	9,70
100000	24.988,80	12.499,00	3.123,70	3.125,00	10,70	10,70
200000	49.999,00	24.999,00	6.250,00	6.250,00	11,70	11,70

	Τρόπος Α		Τρόπος Β		Τρόπος Γ	
N	Χρόνος ανά αναζήτηση		Χρόνος ανά αναζήτηση		Χρόνος ανά αναζήτηση	
	Μεγεθος δεδομένων 55 bytes	Μεγεθος δεδομένων 27 bytes	Μεγεθος δεδομένων 55 bytes	Μεγεθος δεδομένων 27 bytes	Μεγεθος δεδομένων 55 bytes	Μεγεθος δεδομένων 27 bytes
50	10.505,30	12.000,80	16.898,10	18.375,90	4.998,90	4.495,20
100	20.885,70	22.178,20	35.744,40	45.509,90	5.072,60	6.353,20
200	49.672,70	41.102,70	71.875,00	71.220,70	7.289,40	5.047,20

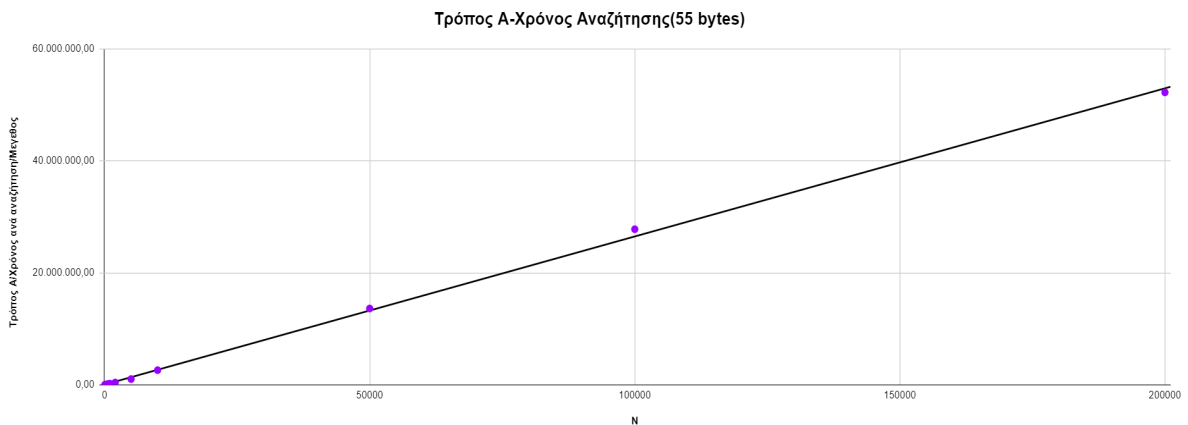
500	106.995,50	103.965,30	148.991,70	149.233,80	4.969,70	4.994,20
800	157.999,70	159.990,30	250.567,10	247.001,90	4.999,80	4.997,80
1000	205.038,40	205.021,90	313.457,80	301.028,00	5.987,70	4.999,70
2000	401.999,40	414.818,00	621.921,20	649.183,40	6.918,40	6.999,90
5000	1.022.792,70	1.029.779,60	1.609.516,30	1.531.406,50	6.999,80	7.998,10
10000	2.615.001,80	3.000.338,50	3.915.926,50	3.030.610,20	11.072,40	21.000,00
50000	13.630.219,5 0	14.057.648,1 0	18.897.739,5 0	18.832.079,1 0	13.323,90	12.358,50
100000	27.796.990,5 0	27.360.793,5 0	38.236.934,8 0	39.091.561,0 0	14.414,30	16.420,80
200000	52.243.622,7 0	53.636.509,8 0	72.752.066,2 0	74.144.682,2 0	21.918,60	16.145,60

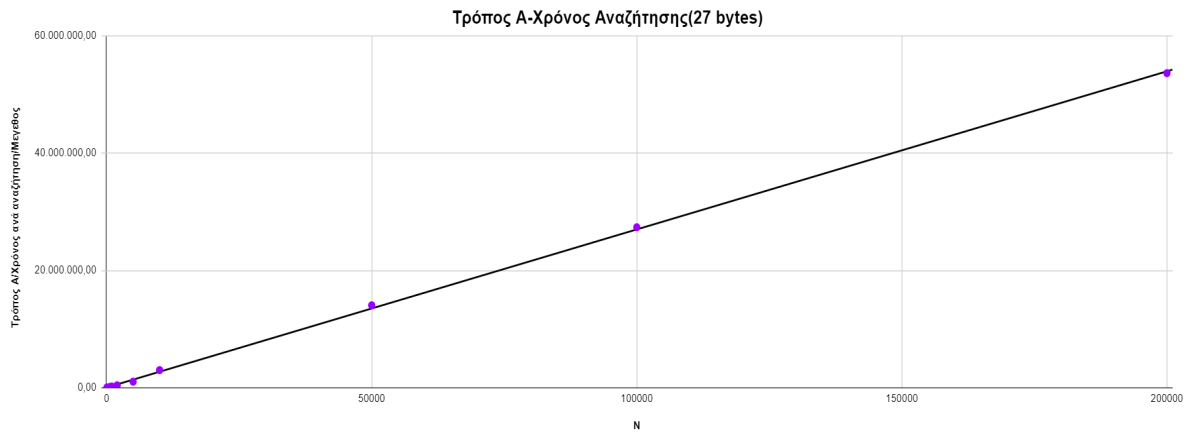
Γραφήματα **Τρόπος Α**





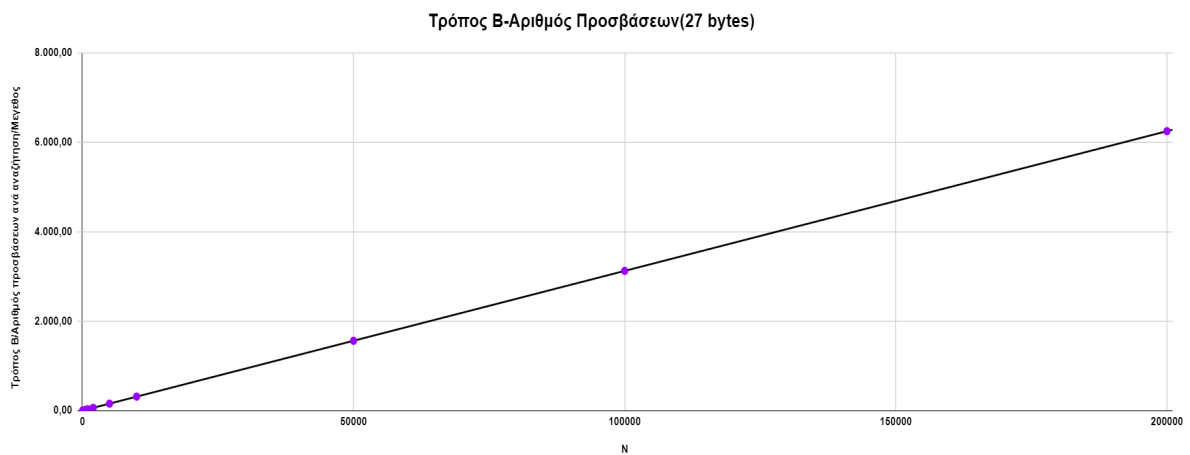
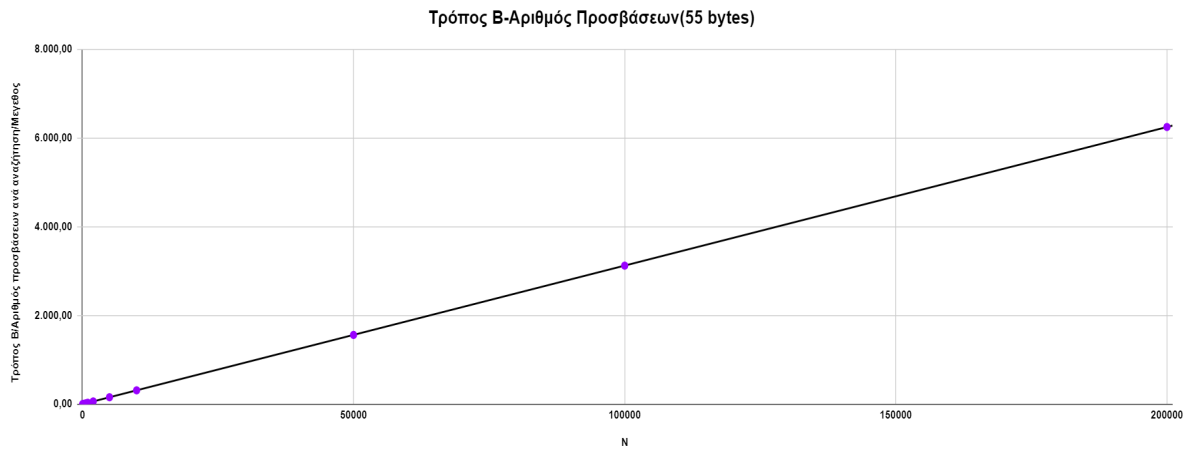
Η μορφή των καμπύλων είναι γραμμική όπως φαίνεται και από τη γραμμή τάσης του διαγράμματος. Γεγονός που μπορούμε να επιβεβαιώσουμε, αφού μετρώντας σειριακά το αρχείο θα πρέπει να αυξάνεται μαζί με το πλήθος των εγγραφών και το μέγεθος του αριθμού προσβάσεων, όπως και γίνεται. Ακόμα αντιλαμβανόμαστε ότι αφού το μέγεθος των εγγραφών στην περίπτωση του string των 27 bytes είναι μικρότερο θα χωράνε περισσότερες εγγραφές σε μια σελίδα άρα θα έχουμε λιγότερες προσβάσεις , εφόσον αυτές γίνονται ανά σελίδα και στο κύριο αρχείο δεδομένων.



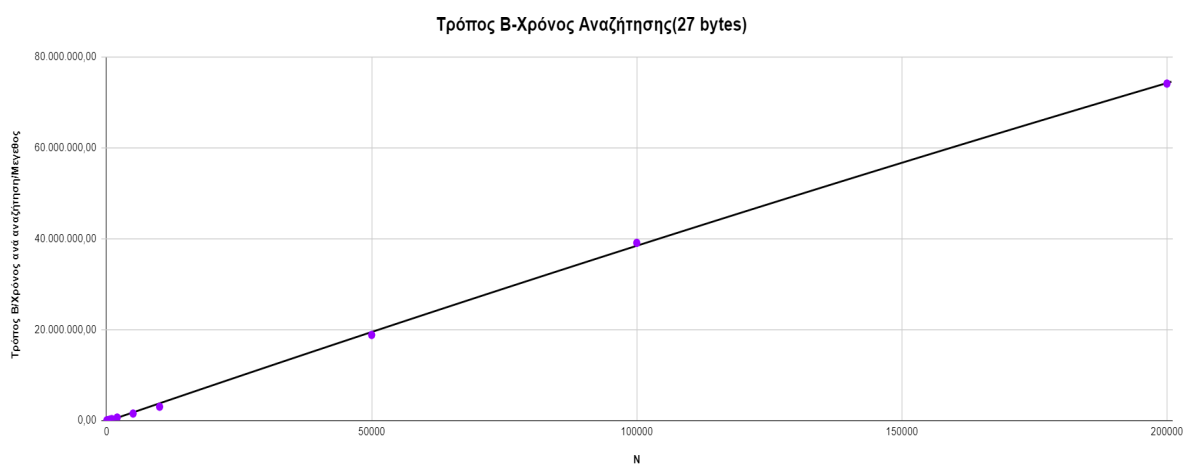
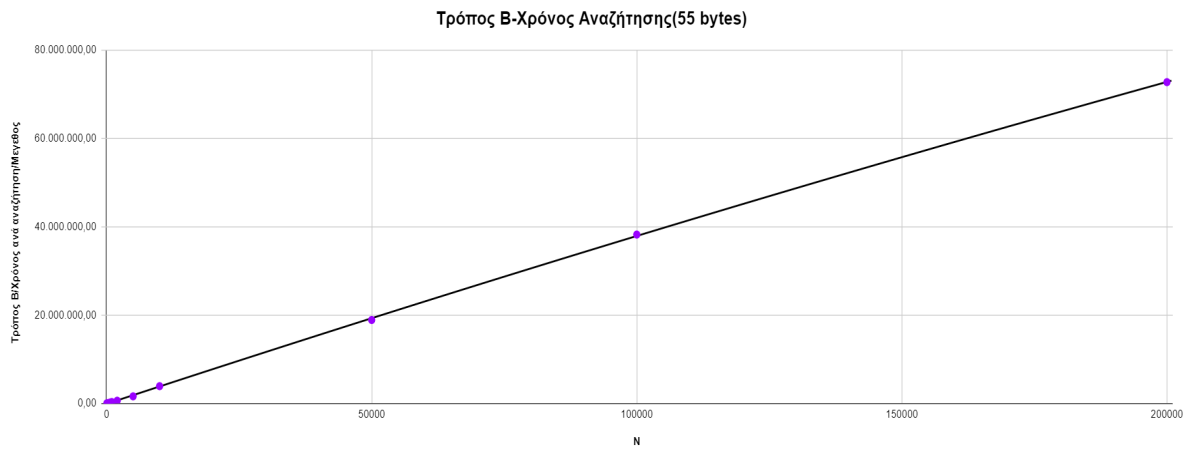


Η μορφή των καμπυλών είναι γραμμική. Αυτό μπορούμε να το επιβεβαιώσουμε από το γεγονός ότι εφόσον η αναζήτηση γινόταν σειριακά το μέγεθος των απαιτούμενων αναζητήσεων αναπόφευκτα θα αυξανώταν μαζί με το μέγεθος των εγγραφών, άρα εν τέλη θα χρειαζόταν και πιο πολύ χρόνος για να ολοκληρωθεί η διαδικασία όσο οι εγγραφές αυξανόντουσαν.

Τρόπος Β

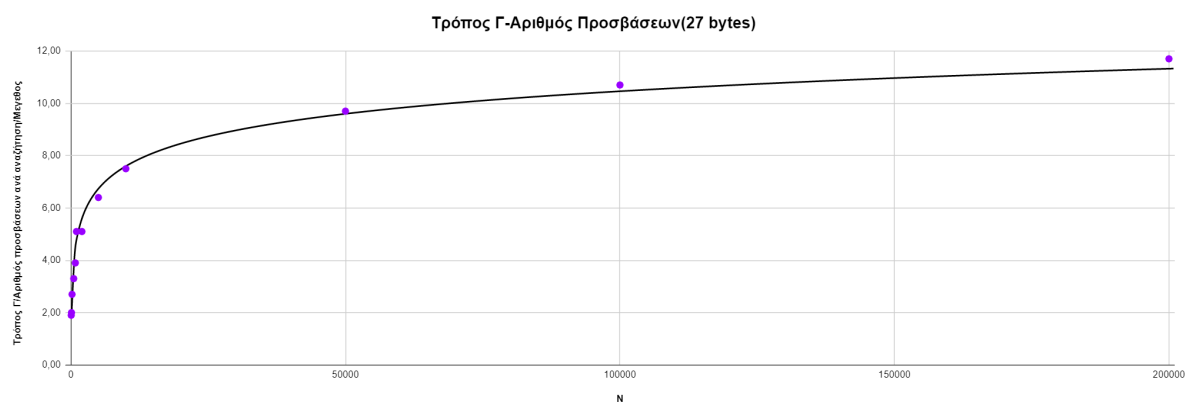
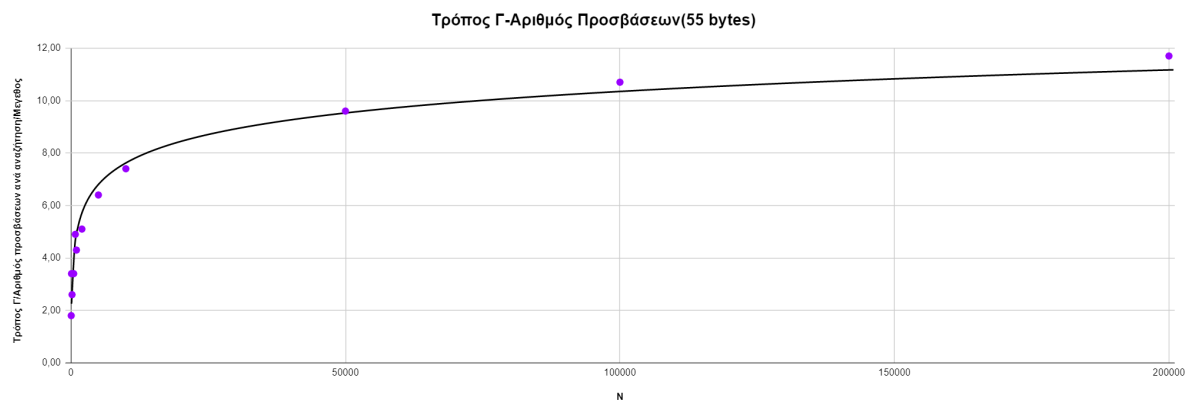


Η μορφή των καμπυλών είναι γραμμική και εξηγείται από το γεγονός πως όσο αυξάνονται οι εγγραφές και η αναζήτηση συνεχίζει να γίνεται σειριακά θα αυξάνεται και ο αριθμός των αναζητήσεων. Βέβαια έχει πολύ μικρότερη κλίση και μεγαλύτερη αποτελεσματικότητα σε σύγκριση με αυτή του Τρόπου Α, αφού το μεγαλύτερο μέρος των προσβάσεων γίνεται στο αρχείο δεικτοδότησης στο οποίο το μέγεθος των εγγραφών είναι πολύ μικρότερο από αυτό του κυρίως αρχείου (8 έναντι 59 ή 31). Για τον λόγο αυτό δεν παρατηρούμε και μεγάλη διαφορά ανάμεσα στις μετρήσεις των 2 μεγεθών.

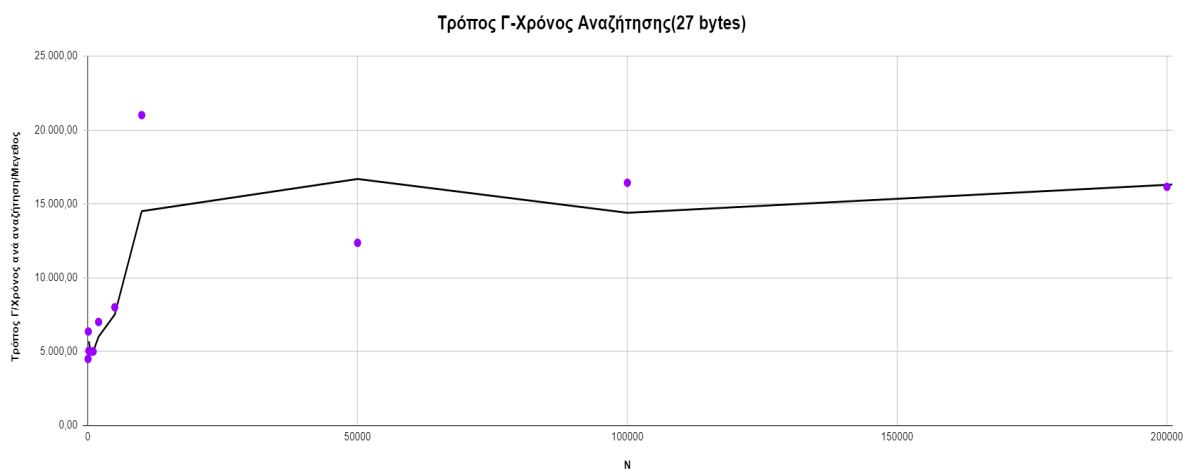
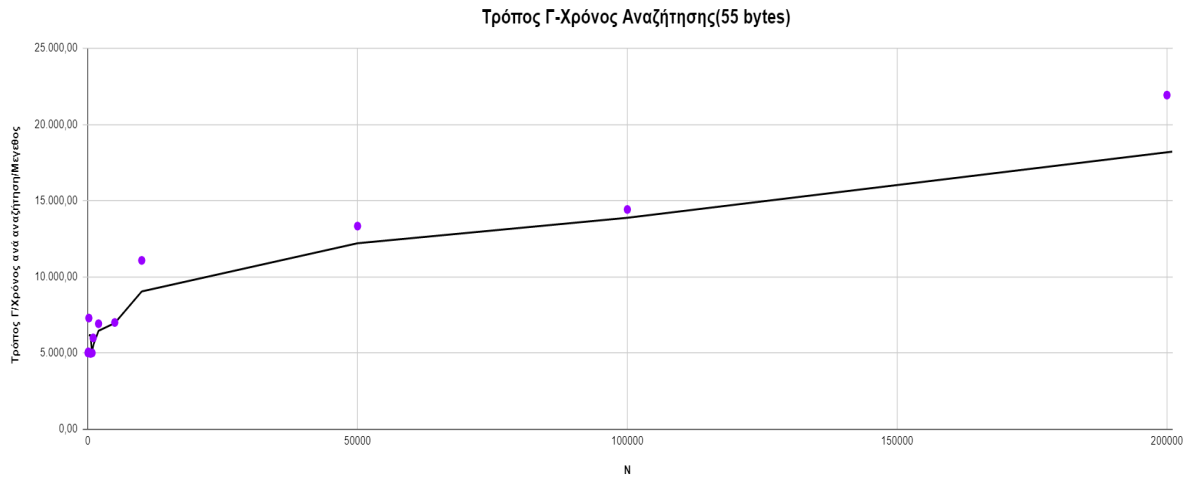


Η καμπύλη του χρόνου αναζήτησης είναι επίσης γραμμική ακριβώς για τον ίδιο λόγο με εκείνον του Τρόπου Α. Επιπροσθέτως, παρατηρούμε ότι γίνεται σε μικρότερο χρονικό διάστημα σε σύγκριση με τον Τρόπο Α λόγω της διαφοράς που προανέφερα στο αρχείο στο οποίο γίνονται οι αναζητήσεις.

Τρόπος Γ



Η μορφή των καμπυλών είναι λογαριθμική όπως φαίνεται από την μορφή της καμπύλης τάσης του διαγράμματος. Αυτό εξηγείται από το γεγονός ότι λόγω της δυαδικής αναζήτησης και το μέγεθος του αρχείου για τις μικρότερες τιμές του N δεν θα υπάρχει μεγάλη μεταβολή αν και φαινομενικά είναι απότομη, λόγω της κλίμακας του διαγράμματος, καθώς είναι πολύ πιο αποτελεσματικός τρόπος από τους άλλους δύο, ενώ όταν θα ξεκινήσει να αυξάνεται το N πάλι δεν θα δούμε ακραία μεταβολή στο μέγεθος του αριθμού προσβάσεων οπότε και η καμπύλη θα μειώσει τον ρυθμό της ανοδικής της πορείας. Η αποτελεσματικότητα του τρόπου αυτού οφείλεται κυρίως στη ταξινόμηση του αρχείου και στο ότι γίνονται πολύ μεγαλύτερα “βήματα” μέσα στο αρχείο που λόγω της ταξινόμησης είναι προς την σωστή κατεύθυνση. Ακόμα δεν παρατηρούμε μεγάλες διαφορές ανάμεσα στα δύο μεγέθη λόγω του ότι η αναζήτηση συνεχίζει να γίνεται στο αρχείο δεικτοδότησης.



Η μορφή των καμπυλών είναι ένας κινούμενος μέσος όρος ο οποίος εξαρτάται από την μεταβολή των τιμών του αριθμού προσβάσεων στον δίσκο. Σε αυτή τη περίπτωση παρατηρούμε ότι ο Τρόπος Γ εκτός από πιο αποτελεσματικός είναι και πιο γρήγορος κυρίως λόγω των περιορισμένων προσβάσεων του.

Γενικά, σε όλους τους τρόπους παρατηρείται εξάρτηση ανάμεσα στον χρόνο αναζήτησης και στον αριθμό προσβάσεων ανάλογη με την μεταβολή των τιμών του, γεγονός που δικαιολογείται και από τη δομή του κώδικα, αφού κάθε φορά που βρίσκεται το ζητούμενο αποτέλεσμα τερματίζει ο βρόχος του οποίου το χρονικό διάστημα εκτέλεσης μετράται για να καταγραφεί ο χρόνος.

4. Ειδικές οδηγίες - Λάθη

Δεν έχω εντοπίσει κάτι συγκεκριμένο στη λειτουργικότητα το κώδικα σίγουρα όμως θα μπορούσε να είναι λίγο μικρότερος. Ειδικές οδηγίες δεν υπάρχουν.