

Γιαννης Μπάρζας 2765

Μερος 1 εξηγηση

Γιαννης Μπάρζας 2765

Βοηθητικες συναρτησεις :

nextValid (sex) : με βαση αν το (sex)=1 για αντρες η -1 για γυναικιες επιστρεφει τα id,age,weight για την επομενη εγκυρη γραμμη.

isValid(age,status): επιστρεφει true αν η γραμμη είναι εγκυρη.

Δομες:

male_dict={}#hash for male

female_dict={}#hash for female

female_max=0 #max value of female

male_max=0 #max value of male

female_cur=0 #current value of female

male_cur=0 #current value of male

heap=[] # the heap!

sum_male_valid=0 # sum all valid lines that A_topk reads for every file

sum_female_valid=0

Αλγοριθμος A_top_k_A() :

Αρχικα βρισκω τις μεγαστες τιμες για τα 2 αρχεια ενημερωνω τις μεταβλητες max,cur και εισαγω τις max γραμμες στα αντιστοιχα hash

Στην συνεχεια υπολογιζω το Τα και ελεγχοντας την καταλληλη συνθηκη κανω το πρωτο pop από τη στοιβα με yield

Μετα μεσα σε ένα φορ που είναι παντα True εχω ένα if αναλογα αν είναι η σειρα του αντρα η της γυναικας και αναλογα την την σειρα περνω την επομενη εγκυρη γραμμη από το καταλληλο αρχαιο

προσθετω(απαραιτητα πεδια) την γραμμη στο καταλληλο hash
ελεγχοντας αν υπαρχει ηδη η age.Ενημερωνω καταλληλα το
current(male_cur h female_cur) και ελενω αν υπαρχει η age στο hash

Του αλλου φυλου και αν υπαρχει δημιουργω ζευγαρια και τα εισαγω
στην στοιβα με το αθροισμα των βαρων πολλαπλασιασμενο με -1 γιατι
η heapq της pythοn είναι min heap.Κρατα τη μεταβλητη turn=turn*(-1)
και την εναλαζω κάθε φορα υπολογιζω την τιμη του Τα και με την
καταλληλη συνθηκη σε ένα if κανω pop το ζευγαρι μεσω yield.

Τελος η συναρτηση ειναι **generator function** και καλειται με την
next.

Εξηγηση μερος 2:

Βοηθητικη συναρτηση

isValid(age,status): επιστρεφει true αν η γραμμη είναι εγκυρη.

Δομες:

male_dict={}#hash for male

heap=[] # the heap!

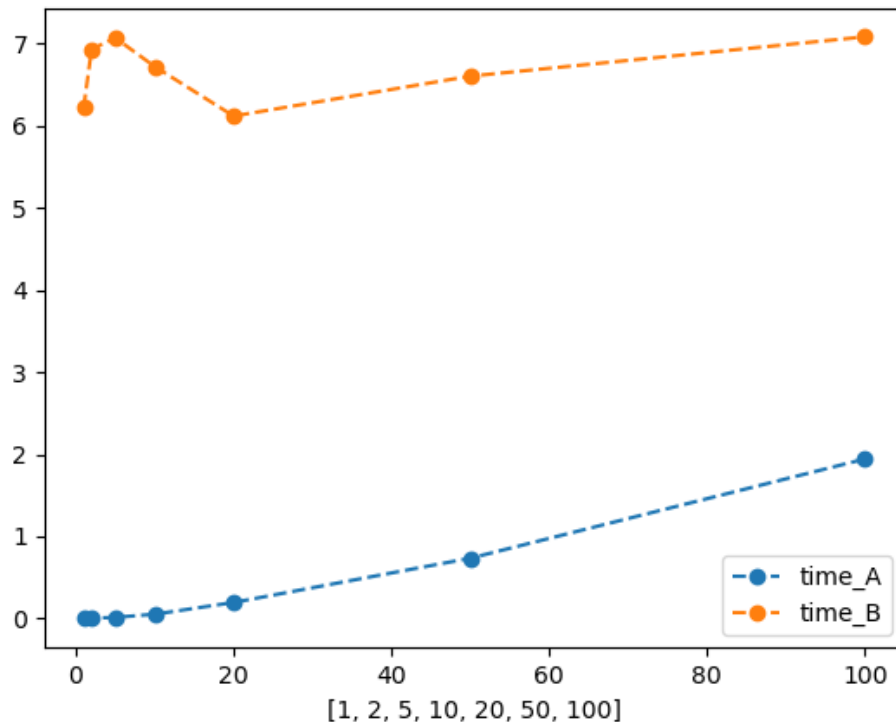
Αλγοριθμος B:

Διαβαζει ολο το αρχειο για τους αντρες και εισαγει τις εγκυρες γραμμες
στο hash male_dict ελεγχοντας αν το age υπαρχει ηδη στη δομη.

Διαβαζει το αρχειο για τις γυναικιες γραμμη γραμμη και αν είναι
εγκυρες ελεχνει αν υπαρχει η ηλικια τηςεγκυρης γραμμης στο
male_dict και αν ναι κανει ζευγαρια και τα εισαγει στην στοιβα αν το
μεγεθος της είναι μικροτερο από κ δηλαδη αν δεν εχει βρει ακομα κ
ζευγαρια αλλιως αν εχει ηδη μεγεθος κ αντικαθιστα το μικροτερο με το
ζευγαρι που εφτιαξε τωρα αν το μικροτερο στοιχειο στην στοιβα είναι
μικροτερο από το ζευγαρι που παει να εισαγει.Σε κάθε κληση βαζει το
μεγαλυτερο στοιχειο στην θεση 0 και κανει pop.

**Η εκτελεση και των 2 προγραμματος γινεται από την
γραμμη εντολων με ορισμα το κ.**

Μερος 3



Σε αυτό το διαγραμμα φαίνεται πως μεταβαляется ο χρονος εκτελεσης των αλγοριθμων A,B συναρτηση της τιμης του κ. Με χρωμα μπλε είναι για τον αλγοριθμο A φαίνεται ότι είναι πιο αποδοτικός από τον αλγοριθμο B καθώς για τις μικρες τιμες του κ ξεκινάει με χρονο κοντα στα 0 δευτερολεπτα και καθώς αυξάνεται ραγδαια το κ ,αυξάνεται γνησια ελαφρως ο χρονος του με μεγιστη τιμη για κ=100 κοντα στα 2 δευτερολεπτα.Αντιθετα ο αλγοριθμος B με πορτοκαλι χρωμα κανει σχεδον σταθερο χρονο που κυμαίνεται στα 6,5 με 7,5 δευτερολεπτα με μια αυξηση στις πρωτες τιμες του κ μια μικρη μειωση που ακολουθειται από μια μικρη σταθερη αυξηση .Η εκτελεση του B είναι σταθερη γιατι διαβαζει εξολοκληρου και τα 2 αρχεια σε αντιθεση με τον αλγοριθμο A ο οποιος εκμεταλευται την ταξινομηση των αρχειων και δεν διαβαζει ολοκληρα τα αρχεια αλλα μεχρι να βρει τα κ κορυφαια ζευγαρια.Οποτε ο A αλγοριθμος εκμεταλευται την ταξινομηση των

αρχειων για να μην τα διαβαζει ολοκληρα τα αρχεια αν δεν είναι απαιτητο ενώ ο B τα διαβαζει όλα ασχετα με τον αριθμο των τοπ ζευγαριων που ψαχνει.

Ο αλγοριθμος B δουλευει σωστα και για μη ταξινομημενα αρχεια ως προ το βαρος ενώ ο αλγοριθμος A δεν μπορεί να δουλεψει σωστα για μη ταξινομημενα αρχεια.Επισης ο αλγοριθμος B κρατα στο male_dict ολο το αρχαιο των αντρων που είναι αρκετα απαιτητικο σε μνημη ενώ ο αλγοριθμος B κρατα μερικες μεταβλητες και 2 λεξικα με επιλεγμενες γραμμες από κάθε αρχαιο αφου εκμεταλευεται την ταξινομηση.

Παρακατω βλεπουμε την **μεταβολη των εγκυρων γραμμων** που διαβαζει ο αλγοριθμος A συναρτηση του κ.Παρατηρουμε μια ραγδαια αυξηση με μεγαστη τιμη για κ=100 τα 10139.Αυτη η αυξηση είναι λογικη γιατι αναζηται περισσοτερα τοπ ζευγαρια . (και για τα 2 αρχεια ιδιος ο αριθμος των αγκυρων γραμμων που διαβαζει ο λογος που δεν φαινεται η δευτερη γραμμη είναι ότι συμπιπτουν).

