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 ιd,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 Α top\_k\_A() :**

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

Στην συνεχεια υπολογιζω το Τα και ελενχοντας 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 αρχειο προσθετω(απαραιτητα πεδια) την γραμμη στο καταλληλο hash ελεχωντας 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 δημιοργω ζευγαρια και τα εισαγω στην στοιβα με το αθροισμα των βαρων πολλαπασιασμενο με -1 γιατι

η heapq της python 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 ποπ 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!

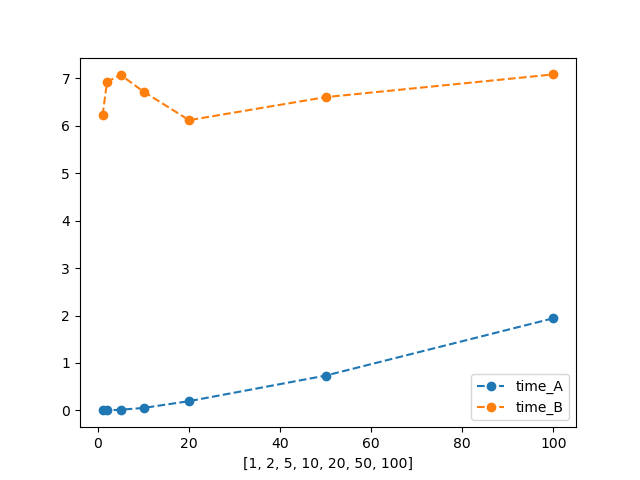
**Αλγοριθμος Β:**

Διαβαζει ολο το αρχειο για τους αντρες και εισαγει τις εγκυρες γραμμες sto hash male\_dict ελενχοντας αν το age υπαρχει ηδη στη δομη.

Διαβαζει το αρχειο για τις γυναικιες γραμμη γραμμη και αν είναι εγκυρες ελεχνει αν υπαρχει η ηλικεια τηςεγκυρης γραμμης στο male\_dict και αν ναι κανει ζευγαρια και τα εισαγει στην στοιβα αν το μεγεθως της είναι μικροτερο από κ δηλαδη αν δεν εχει βρει ακομα κ ζευγαρια αλλιως αν εχει ηδη μεγεθως 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.

**H εκτελεση και των 2 προγραμματων γινεται από την γραμμη εντολων με ορισμα το κ.**

**Μερος 3**



Σε αυτό το διαγραμμα φαινεται πως μεταβαλεται ο χρονος εκτελεσης ταν αλγοριθμων Α,Β συναρτηση της τιμης του κ. Με χρωμα μπλε είναι για τον αλγοριθμο Α φαινεται ότι είναι πιο αποδοτικος από τον αλγοριθμο Β καθως για τις μικρες τιμες του κ ξεκιναει με χρονο κοντα στα 0 δευτερολεπτα και καθως αυξανεται ραγδαια το κ ,αυξανεται γνησια ελαφρως ο χρονος του με μεγιστη τιμη για κ=100 κοντα στα 2 δευτερολεπτα.Αντιθετα ο αλγοριθμος Β 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 Β είναι σταθερη γιατι διαβαζει εξολοκληρου και τα 2 αρχεια σε αντιθεση με τον αλγοριθμο Α ο οποιος εκμεταλευται την ταξινομηση των αρχειων και δεν διαβαζει ολοκληρα τα αρχεια αλλα μεχρι να βρει τα κ κορυφαια ζευγαρια.Οποτε ο Α αλγοριθμος εκμεταλευται την ταξινομηση των αρχειων για να μην τα διαβαζει ολοκληρα τα αρχεια αν δεν είναι απαραιτητο ενώ ο Β τα διαβαζει όλα ασχετα με τον αριθμο των τοπ ζευγαριων που ψαχνει.

Ο αλγοριθμος Β δουλευει σωστα και για μη ταξινομημενα αρχεια ως προ το βαρος ενώ ο αλγοριθμος Α δεν μπορει να δουλεψει σωστα για μη ταξινομημενα αρχεια.Επισης ο αλγοριθμος Β κρατα στο male\_dict ολο το αρχειο των αντρων που είναι αρκετα απαιτητικο σε μνημη ενώ ο αλγοριθμος Β κρατα μερικες μεταβλητες και 2 λεξικα με επιλεγμενες γραμμες από κάθε αρχειο αφου εκμεταλευεται την ταξινομηση.

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

