

## ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ Η/Υ, ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ Ι, ΑΚΑΔΗΜΑΪΚΟ ΕΤΟΣ 2018-2019

ΕΡΓΑΣΙΑ ΕΞΑΜΗΝΟΥ

### ***Ποιος πρέπει να ολοκληρώσει αυτή την εργασία?***

Φοιτητές έτους  $\geq 2$  που δεν πέρασαν επιτυχώς τα εργαστήρια ή την εργασία εξαμήνου του ακαδημαϊκού έτους 2017-2018 και επιθυμούν να συμμετέχουν στην εργαστηριακή εξέταση Ιανουαρίου ή/και Σεπτεμβρίου. Για περισσότερες λεπτομέρειες δείτε τη σελίδα του μαθήματος.

Η εργασία μπορεί να γίνει ατομικά ή σε ομάδες των 2 ατόμων.

### ***Διαδικαστικά***

Προθεσμία: 07/1/2019, 23:59

Η προθεσμία υποβολής είναι οριστική. Εκπρόθεσμες ασκήσεις δε γίνονται δεκτές.

Οδηγίες για την υποβολή της άσκησης θα αναρτηθούν αργότερα.

## Γενική περιγραφή

Θα υλοποιήσετε μια εφαρμογή αξιολόγησης κριτικών για ταινίες. Η εφαρμογή σας θα μελετήσει μια συλλογή κριτικών οι οποίες έχουν ήδη αξιολογηθεί και με βάση τα στοιχεία που θα συλλέξει θα μπορεί να αξιολογήσει νέες κριτικές που εισάγονται από το πληκτρολόγιο.

Τα δεδομένα που θα χρησιμοποιήσει η εφαρμογή σας είναι μια συλλογή από μερικές χιλιάδες κριτικές ταινιών κάθε μία εκ των οποίων έχει ήδη αξιολογηθεί με έναν ακέραιο 0-4 με σημασία:

- 0 - αρνητική
- 1 - κάπως αρνητική
- 2 - ουδέτερη
- 3 - κάπως θετική
- 4 - θετική

Τα δεδομένα παρέχονται σε αρχείο κειμένου. Κάθε γραμμή του αρχείου περιέχει το βαθμό αξιολόγησης μιας κριτικής και ακολούθως το κείμενο αυτής της κριτικής. Θα σας δώσουμε μια έτοιμη συνάρτηση η οποία σε κάθε κλήση της αποθηκεύει σε ένα πίνακα χαρακτήρων την επόμενη γραμμή του αρχείου.

Το πρόγραμμά σας θα πρέπει να αποσπά τις λέξεις κάθε κριτικής και να τις αποθηκεύει σε μια δομή δεδομένων που θα σας περιγράψουμε, μαζί με ένα αθροιστικό σκορ που προκύπτει από το βαθμό αξιολόγησης της κριτικής.

Αφού συλλεχθούν όλα τα δεδομένα, το πρόγραμμα παρουσιάζει στο χρήστη ένα μενού επιλογών για διάφορες επιθυμητές λειτουργίες.

**Προσοχή: Διαβάστε ΟΛΗ την εκφώνηση πριν αρχίσετε να γράφετε κώδικα.**

## Απαιτήσεις

Απαγορεύεται αυστηρά η χρήση καθολικών μεταβλητών.

Απαγορεύεται αυστηρά η χρήση goto.

Κάθε καλά ορισμένη λειτουργία (πχ. αναζήτηση) πρέπει να γίνεται μέσω κατάλληλης συνάρτησης που έχετε εσείς ορίσει.

Ορίστε τα μεγέθη των πινάκων με #define και γράψτε τον κώδικά σας έτσι ώστε αν αλλαχθεί κάποιο μέγεθος, να συνεχίσει να λειτουργεί σωστά το πρόγραμμα.

Επειδή ο στόχος της εργασίας είναι να εξασκηθείτε σε διάφορες δομές και στοιχεία της γλώσσας, κάποιες από τις λειτουργίες πρέπει να υλοποιηθούν με το συγκεκριμένο τρόπο που σας περιγράφουμε.

Χρησιμοποιήστε τα ενδεικτικά αρχεία εισόδου/εξόδου που θα σας δώσουμε για να ελέγξετε την ορθότητα της εφαρμογής σας. Η έξοδός σας πρέπει να ταιριάζει ακριβώς με τη δική μας.

## Δομές δεδομένων

Η βασική δομή δεδομένων είναι ένας **πίνακας εγγραφών** μεγέθους 20.000 στον οποίο θα αποθηκευτούν οι λέξεις που εμφανίζονται στις κριτικές μαζί με τα αθροιστικά σκορ αξιολόγησης της κάθε μίας. Για την ακρίβεια, κάθε κελί του πίνακα περιέχει :

- μια λέξη ως πίνακα χαρακτήρων μεγέθους 30
- το πλήθος εμφανίσεων της λέξης σε κριτικές.
- το αθροιστικό σκορ της λέξης (άθροισμα των σκορ των κριτικών στις οποίες εμφανίζεται η λέξη).

Από τα δύο τελευταία πεδία μπορεί να προκύψει το μέσο σκορ αξιολόγησης μιας λέξης.

Τα περιεχόμενα του πίνακα εγγραφών πρέπει να διατηρούνται ταξινομημένα σε λεξικογραφικά αύξουσα σειρά με βάση τη λέξη που περιέχει κάθε κελί. Συνίσταται να υλοποιήσετε ταξινομημένη εισαγωγή εγγραφής και όχι απλή εισαγωγή και μετά ταξινόμηση.

Επιπλέον, θα κατασκευάσετε ένα **πίνακα δεικτών** μεγέθους 27 όπου κάθε κελί του από 0 έως και 25 περιέχει τη διεύθυνση του πρώτου κελιού του πίνακα εγγραφών για τις λέξεις που ξεκινούν από το αντίστοιχο γράμμα του αγγλικού αλφαβήτου. Με άλλα λόγια, ο δείκτης στη θέση 0 του πίνακα δεικτών θα δείχνει στο κελί του πίνακα εγγραφών όπου βρίσκεται η πρώτη λέξη που ξεκινά από 'Α'. Ο δείκτης στη θέση 1 του πίνακα δεικτών θα δείχνει στο κελί του πίνακα εγγραφών όπου βρίσκεται η πρώτη λέξη που ξεκινά από 'Β', κ.ο.κ. Αν δεν υπάρχουν λέξεις που να ξεκινούν από ένα συγκεκριμένο γράμμα, τότε ο αντίστοιχος δείκτης είναι NULL. Επιπλέον, στην τελευταία θέση του πίνακα δεικτών υπάρχει δείκτης που δείχνει πάντα στην τελευταία έγκυρη εγγραφή του πίνακα εγγραφών.

## Λειτουργία προγράμματος

Ακολουθούν λεπτομέρειες για το πώς ακριβώς πρέπει να λειτουργεί το πρόγραμμά σας.

### Επεξεργασία γραμμής εντολής

Ο χρήστης προσδιορίζει το όνομα ενός αρχείου δεδομένων (κριτικών) στη γραμμή εντολής.

Αν δεν έχει δοθεί σωστό πλήθος παραμέτρων στη γραμμή εντολής, το πρόγραμμα τερματίζει αφού εκτυπώσει το μήνυμα **Incorrect number of parameters.** ακολουθούμενο από **χαρακτήρα αλλαγής γραμμής**.

### Εισαγωγή δεδομένων και κατασκευή δομών

*Κατασκευή πίνακα εγγραφών.*

Σας δίνουμε τη συνάρτηση `int getNextLine(char filename[], char line[], int maxlinesize)` η οποία παίρνει ως παραμέτρους το όνομα του αρχείου δεδομένων, έναν αρχικά άδειο πίνακα χαρακτήρων και το μέγεθος του πίνακα `line`. Κάθε φορά που καλείται η συνάρτηση, αν υπάρχει επόμενη γραμμή στο αρχείο, την αποθηκεύει στον πίνακα `line` και επιστρέφει 1. Αν δεν υπάρχει άλλη γραμμή στο αρχείο, επιστρέφει 0.

Γράψτε μια συνάρτηση που παίρνει ως παράμετρο ένα αρχικά άδειο πίνακα εγγραφών και ότι άλλο χρειάζεται, και η οποία για κάθε γραμμή που διαβάζεται από την κλήση της `getNextLine`:

- αποσπά την πρώτη "λέξη" η οποία είναι πάντα μια ακέραια τιμή που εκφράζει το σκορ αξιολόγησης της κριτικής.
- αποσπά μία-μία τις επόμενες λέξεις. Κάθε μία από αυτές τη μετατρέπει σε μικρά γράμματα και, αν βρίσκεται ήδη στον πίνακα εγγραφών, ανανεώνει το το πλήθος εμφανίσεων της καθώς και το αθροιστικό της σκορ προσθέτοντας σε αυτό το σκορ που μόλις διαβάστηκε. Αν δε βρίσκεται, εισάγεται με πλήθος 1 και αρχικό αθροιστικό σκορ αυτό που μόλις διαβάστηκε.

Οι λέξεις διαχωρίζονται με χαρακτήρες `tab` ή/και `space`. Τα σημεία στίξης αντιμετωπίζονται ως κανονικές λέξεις.

### Κατασκευή πίνακα δεικτών.

Γράψτε μια συνάρτηση η οποία παίρνει ως παραμέτρους τον πίνακα εγγραφών, έναν αρχικά άδειο πίνακα δεικτών προς εγγραφές και ότι άλλο χρειάζεται και αρχικοποιεί τον πίνακα δεικτών όπως περιγράφεται στην ενότητα "Δομές δεδομένων".

### Μενού

Αφού κατασκευαστούν οι δομές, το πρόγραμμα σε επανάληψη εκτυπώνει το παρακάτω μενού και, ανάλογα με την επιλογή του χρήστη, εκτελεί την αντίστοιχη λειτουργία.

1. Print group of words
2. Print score of word
3. Best score
4. Worst score
5. Evaluate review
6. Exit

Ακολουθεί χαρακτήρας αλλαγής γραμμής.

Η επιλογή 6 τερματίζει την επανάληψη και το πρόγραμμα. Οι υπόλοιπες επιλογές αναλύονται παρακάτω.

### Επιλογή 1: Print group of words

Γράψτε μια βοηθητική συνάρτηση η οποία παίρνει ως παραμέτρους τον πίνακα δεικτών και ένα γράμμα και, μέσω του πίνακα δεικτών, εντοπίζει κι εκτυπώνει όλες τις εγγραφές που περιέχουν λέξεις που ξεκινούν από αυτό το γράμμα. Για την ακρίβεια, για κάθε λέξη που βρίσκει, εκτυπώνει:

- το πλήθος εμφανίσεων αυτής με πλάτος 2 χαρακτήρων
- ένα κενό (space)
- το αθροιστικό της σκορ της με ένα δεκαδικό ψηφίο και συνολικό πλάτος 4 χαρακτήρων
- ένα κενό (space)
- τη λέξη ανάμεσα σε χαρακτήρες "
- χαρακτήρα αλλαγής γραμμής

Αν δε βρεθούν λέξεις από αυτό το γράμμα, τότε εκτυπώνει το μήνυμα **No words.** ακολουθούμενο από χαρακτήρα αλλαγής γραμμής.

Στην επιλογή 1, το πρόγραμμα εκτυπώνει το μήνυμα **Enter starting letter:** (υπάρχει ένα space μετά την άνω-κάτω τελεία) και διαβάζει από το πληκτρολόγιο ένα γράμμα το οποίο μπορεί να είναι κεφαλαίο ή μικρό. Αν διαβαστεί κάτι που δεν είναι γράμμα του αγγλικού αλφαβήτου, εκτυπώνει το μήνυμα **Invalid letter.** ακολουθούμενο από χαρακτήρα αλλαγής γραμμής. Διαφορετικά, καλεί τη βοηθητική συνάρτηση που γράψατε.

### Επιλογή 2: Print score of word

Γράψτε μια βοηθητική συνάρτηση η οποία παίρνει ως παραμέτρους τον πίνακα εγγραφών, μία λέξη και ότι άλλο χρειάζεται και χρησιμοποιεί δυαδική αναζήτηση ώστε να εντοπίσει αυτή τη λέξη στον πίνακα εγγραφών. Αν τη βρει, επιστρέφει τη θέση της στον πίνακα, διαφορετικά επιστρέφει -1.

Στην επιλογή 2, το πρόγραμμα εκτυπώνει το μήνυμα **Enter word:** (υπάρχει ένα space μετά την άνω-κάτω τελεία) και διαβάζει από το πληκτρολόγιο μία λέξη. Ακολούθως καλεί τη βοηθητική συνάρτηση που γράψατε. Αν η λέξη υπάρχει, τότε εκτυπώνει το μήνυμα **Score: X.XX.** όπου X.XX το μέσο σκορ της λέξης (όπως αυτό προκύπτει από το αθροιστικό σκορ και το πλήθος εμφανίσεων αυτής σε κριτικές). Το μέσο σκορ πρέπει να εμφανίζεται με δύο δεκαδικά ψηφία και να ακολουθείται από χαρακτήρα αλλαγής γραμμής. Αν η λέξη δεν υπάρχει στον πίνακα, το πρόγραμμα εκτυπώνει το μήνυμα **No score for "W"** όπου W η λέξη και ακολουθεί χαρακτήρας αλλαγής γραμμής.

**Επιλογές 3, 4: Best/Worst score.**

Γράψτε μια βοηθητική συνάρτηση η οποία παίρνει ως παραμέτρους τον πίνακα εγγραφών, το πλήθος εγγραφών και μια ένδειξη για το αν αναζητάμε το καλύτερο ή το χειρότερο σκορ και **με χρήση αναδρομής** βρίσκει κι επιστρέφει την εγγραφή που περιλαμβάνει τη λέξη με το καλύτερο ή το χειρότερο μέσο σκορ.

Στις επιλογές 3 και 4 το πρόγραμμα καλεί τη βοηθητική συνάρτηση που γράψατε με κατάλληλες παραμέτρους και μετά εκτυπώνει το μήνυμα **Word: W, score: X.XX** όπου W η λέξη που περιλαμβάνεται στην εγγραφή που βρέθηκε και X.XX το μέσο σκορ αυτής με δύο δεκαδικά ψηφία. Ακολουθεί **χαρακτήρας αλλαγής γραμμής**. Αν υπάρχουν πολλές λέξεις με ίδιο μέσο σκορ, θα πρέπει να επιλεχθεί όποια εμφανίζεται μεταγενέστερα στον πίνακα.

**Επιλογή 5: Evaluate review**

Γράψτε μια βοηθητική συνάρτηση η οποία παίρνει ως παραμέτρους τον πίνακα εγγραφών, ένα πίνακα που περιέχει μια κριτική, και ότι άλλο χρειάζεται κι εκτυπώνει μια εκτίμηση της κριτικής. Η συνάρτηση υπολογίζει το συνολικό σκορ αξιολόγησης της κριτικής το οποίο είναι το άθροισμα των μέσων σκορ των λέξεων που την αποτελούν δια του πλήθους αυτών.

Αν αυτό είναι μηδέν (π.χ. επειδή καμία από τις λέξεις δε βρίσκεται στον πίνακα εγγραφών), τότε εκτυπώνεται το μήνυμα **This review cannot be evaluated.** ακολουθούμενο από **χαρακτήρα αλλαγής γραμμής**. Διαφορετικά, εκτυπώνεται το μήνυμα **Score: X.XX. This review is** (υπάρχει ένα κενό μετά το is) κι ανάλογα με το αν το σκορ αξιολόγησης είναι μικρότερο, ίσο ή μεγαλύτερο του 2, ακολουθεί η λέξη **negative, neutral** ή **positive** αντίστοιχα. Ακολουθεί **τελεία και χαρακτήρας αλλαγής γραμμής**. Το X.XX είναι το σκορ της κριτικής με δύο δεκαδικά ψηφία.

Το πρόγραμμα εκτυπώνει το μήνυμα **Enter review:** (με ένα κενό μετά το ':'), διαβάζει μια γραμμή που περιέχει μια κριτική και καλεί τη βοηθητική συνάρτηση που γράψατε ώστε να εκτιμήσει την κριτική.