

# ΕΙΣΑΓΩΓΗ ΣΤΟΝ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟ (2009-10)

## Άσκηση 2

Γράψτε ένα πρόγραμμα C (έστω ότι το πηγαίο αρχείο του ονομάζεται `textstats.c`) το οποίο να διαβάζει από την είσοδο ένα κείμενο σε φυσική γλώσσα και κάθε γραμμή που διαβάζει να την εκτυπώνει πάλι στην έξοδο, έχοντας προτάξει στην αρχή της τον αύξοντα αριθμό της.

Στο τέλος, αφού δοθεί όλη η είσοδος, το πρόγραμμά σας να εκτυπώνει τον αύξοντα αριθμό της γραμμής που περιέχει τις περισσότερες λέξεις. Επίσης, να δίνει την πληροφορία ποια λέξη από όλο το κείμενο είναι εκείνη με τους περισσότερους χαρακτήρες, εκτυπώνοντας τον αριθμό γραμμής στην οποία βρίσκεται η λέξη αυτή και τον αύξοντα αριθμό της λέξης μέσα στη γραμμή.

Σαν λέξη θεωρείται μία ακολουθία από χαρακτήρες που δεν περιλαμβάνει κενούς χαρακτήρες, χαρακτήρες στηλογνώμονα (`'\t'`), χαρακτήρες αλλαγής γραμμής (`'\n'`) ή σημεία στίξης. Σαν σημεία στίξης θεωρούνται η τελεία (`'.'`), το κόμμα (`','`), το θαυμαστικό (`'!'`), το αγγλικό ερωτηματικό (`'?'`) και η αγγλική άνω τελεία (`';'`).

Αν υπάρχουν δύο ή περισσότερες γραμμές με τον ίδιο μέγιστο αριθμό λέξεων, το πρόγραμμά σας να εκτυπώνει τον αύξοντα αριθμό μόνο της πρώτης από αυτές. Επίσης, αν υπάρχουν δύο ή περισσότερες λέξεις με τον ίδιο μέγιστο αριθμό χαρακτήρων, το πρόγραμμά σας να εκτυπώνει τα απαιτούμενα στοιχεία (αύξοντα αριθμό γραμμής και αύξοντα αριθμό λέξης μέσα στη γραμμή) μόνο για την πρώτη από αυτές.

Κάποιες ενδεικτικές εκτελέσεις του προγράμματος φαίνονται στη συνέχεια (τα αρχεία που χρησιμοποιούνται μπορούν να βρεθούν στο <http://www.di.uoa.gr/~ip/hwfiles>).

```
% ./textstats
This is a text given from
Line 00001: This is a text given from
the keyboard to test the correctness of
Line 00002: the keyboard to test the correctness of
the program textstats.c. You may check your
Line 00003: the program textstats.c. You may check your
program by giving the same input, as well.
Line 00004: program by giving the same input, as well.
^D
Line 3 with maximum number of words (8)
Word 6 in line 2 with maximum number of characters (11)
```

```
% ./textstats
A line with 5 words
Line 00001: A line with 5 words
One more line with 6 words
Line 00002: One more line with 6 words
And another one with 6 words
Line 00003: And another one with 6 words
And another one
Line 00004: And another one
^D
Line 2 with maximum number of words (6)
Word 2 in line 3 with maximum number of characters (7)
```

```
% ./textstats < demotext.txt
Line 00001: Oh , my God ! ! What
Line 00002: an ugly file , isn't
Line 00003:
Line 00004: it? Yes ,it is ; could
Line 00005: you write a C program
Line 00006:
Line 00007:
Line 00008: to get its statistics ,please
Line 00009: ? Thank you ! !
```

Line 4 with maximum number of words (5)  
Word 4 in line 8 with maximum number of characters (10)

```
% ./textstats < KRexcerpt.txt
Line 00001: C is a general-purpose programming language. It has been closely
Line 00002: associated with the UNIX system where it was developed, since both
Line 00003: the system and most of the programs that run on it are written in
Line 00004: C. The language, however, is not tied to any one operating system
Line 00005: or machine; and although it has been called a "system programming
Line 00006: language" because it useful for writing compilers and operating
Line 00007: systems, it has been used equally well to write major programs in
Line 00008: many different domains.
```

Line 3 with maximum number of words (14)  
Word 4 in line 1 with maximum number of characters (15)

```
% ./textstats < largefile.txt
Line 00001: Informix Messages and Corrections.                Publication Date: Februrary 1999
Line 00002: Version 1.1                                                Message Range: 73018 - 71572
Line 00003: # END OF VERSION INFORMATION
Line 00004:
Line 00005:
Line 00006: 0 Success.
.....
Line 46092: This is the same as error -460 returned by the database server.
Line 46093:
Line 46094:
Line 46095:
Line 46096: -----
```

Line 20777 with maximum number of words (33)  
Word 1 in line 45098 with maximum number of characters (70)

Η παράδοση της άσκησης αυτής συνίσταται στην υποβολή του πηγαίου αρχείου textstats.c με διαδικασία που θα ανακοινωθεί σύντομα.

**Υπόδειξη:** Υλοποιήστε το πρόγραμμά σας σταδιακά. Σε πρώτη φάση ασχοληθείτε μόνο με το να εκτυπώνονται οι γραμμές που διαβάζονται από την είσοδο με τον αύξοντα αριθμό τους στην αρχή. Αφού βεβαιωθείτε ότι αυτό δουλεύει σωστά, ασχοληθείτε με την εύρεση της γραμμής με τις περισσότερες λέξεις και, αφού γίνει και αυτό, στο τέλος δείτε και την εύρεση της λέξης με τους περισσότερους χαρακτήρες.