

Εργαστήριο #5: Ασκήσεις Λεξικής Ανάλυσης

1. Να γίνει λεξικός αναλυτής που αναγνωρίζει τα εξής:

- a. Τα ρήματα

is|am|are|were|was|be|being|been|do|does|did|should|can|could|has|
have|had|go

- b. Τις άλλες λέξεις

- c. Τα κενά διαστήματα (ακολουθίες από κενά και tab)

- d. Οτιδήποτε άλλο

Όταν δίνετε ρήμα εμφανίζει ένα μήνυμα σαν κι αυτό: did: is a verb

Όταν δίνετε άλλη λέξη εμφανίζει ένα μήνυμα σαν κι αυτό: day: is not a verb

Τα κενά διαστήματα τα αγνοεί.

Οτιδήποτε άλλο το αντηχεί στην οθόνη.

Κάνετε δύο παραλλαγές του προγράμματος μία με την είσοδο από το πληκτρολόγιο και μία από αρχείο το όνομα του οποίου δίνεται από την γραμμή εντολών.

2. Στον flex υπάρχουν οι ακόλουθες εκφράσεις κλάσεων:

[:alnum:] [:alpha:] [:blank:]

[:cntrl:] [:digit:] [:graph:]

[:lower:] [:print:] [:punct:]

[:space:] [:upper:] [:xdigit:]

Αυτές πρέπει να εμφανίζονται μέσα στους χαρακτήρες [] που χρησιμοποιούνται για τις κανονικές κλάσεις και μπορούν να συνδυαστούν μεταξύ τους ή και με άλλους χαρακτήρες μέσα σε μία κλάση. Για παράδειγμα η κλάση [[:alpha:][:digit:]] έχει το ίδιο νόημα με την κλάση [[:alpha:]0-9] και με την κλάση [a-zA-Z0-9].

Δημιουργείτε έναν λεξικό αναλυτή που αναγνωρίζει λέξεις με μικρά και κεφαλαία γράμματα καθώς και δεκαεξαδικούς αριθμούς χρησιμοποιώντας τις εκφράσεις [:lower:], [:upper:] και [:xdigit:] αντίστοιχα.

Η έξοδος του προγράμματος πρέπει να είναι όπως φαίνεται παρακάτω:

BIG

Found an uppercase word: BIG

low

Found a lower case word: low

12F

Found a hexadecimal number: 12F

Κάνετε δύο παραλλαγές του προγράμματος μία με την είσοδο από το πληκτρολόγιο και μία από αρχείο το όνομα του οποίου δίνεται από την γραμμή εντολών.

3. Επεκτείνετε την άσκηση 2, ώστε να καταμετρά τις λέξεις με μικρά και κεφαλαία γράμματα, καθώς και τους δεκαδικούς και δεκαεξαδικούς αριθμούς όταν ο λεξικός αναλυτής εκτελείται σε ένα αρχείο.